Funded by
the European Union

ISBN: 978---615-0-2-1-9
9 786150 219790

Digital Individual Learning Accounts in the Visegrad Countries

WP5 - Methodological Guide

Project Code: 2022-1-HU01-ADU-000086794
Action Type: KA220-ADU-Cooperation partnership in adult education

KIFÜ

SGH Warsaw School of Economics

National Training Fund

SIOV štátny inštitút odborného vzdelávania

Impressum

Published by:

Governmental Agency for IT Development - president

Editor: Erzsébet Szlamka (HU)

Proof-reader: Emma Wyatt (SK)

Graphic Design: Teodóra Matisz (HU)

Authors: Zdeňka Šímová (CZ), Markéta Tučková (CZ), Erzsébet Szlamka (HU), György Salomvári (HU), Horacy Dębowski (PL), Wojciech Stęchły (PL), Ľubica Gállová (SK), Zuzana Izáková (SK), Katarína Šmálová (SK)

Special thanks: Réka Vonnák, István Nagy-Rácz, dmlab.hu

Address: Governmental Agency for IT Development 1134 Budapest, Vaci str. 35.

www.kifu.gov.hu

Email: d-ila@kifu.hu

Funded by the European Union

Table of content

I. Introduction 2
II. Conceptual framework of the D-ILA data model 2
II.1 Policy modelling 2
II.2 Data in the D-ILA data model 3
II.3 Individual Learning Accounts (ILAs) 4
II.4 Soft skills and related frameworks 6
II.5 Soft skills frameworks in the EU 8
II.6 Application of Artificial Intelligence in Modelling 16
III. Preparations for the creation of the D-ILA data model 17
III.1 Conclusions on adult learning systems and their funding solutions 17
III.2 Conclusions based on the presentation of national soft skills solutions 18
III.3 General findings on the ILA data model 19
III.4 Recommendations 21
III.5 Findings concerning the development of the ILA data model 23
IV. Introducing the D-ILA data model 27
IV.1 Compilation of data fields 28
IV.2 Compilation of value sets 30
IV.3 Generation of individuals' data 32
IV.4 Generation of training data 38
IV.5 The matching exercise 46
V. Testing the D-ILA data model using AI 48
V.1 Policy scenarios tested using the D-ILA data model 48
V.2 Indicators developed to evaluate the results of AI runs 52
V.3 How AI works 57
V.4 Results of AI runs 59
V.5 Conclusions for the D-ILA data model 63
VI. Methodological guide for the application of the model 65
VI.1 Overview 65
VI.2 Sequence of steps to make your own D-ILA data model 65
VI.3 Products that help to implement the sequence of steps 68
VI.4 Using the D-ILA data model for the evaluation of policy intervention scenarios 71
VI.5 Q&A for the D-ILA data model 72
VII. Use cases for the D-ILA data model 74
VIII. Annex 81
VIII.1 AI Run Reports 81
VIII.2 Commensurability matrix 81

Funded by
the European Union

I. Introduction

Dear reader, you have in your hands a Methodological Guide. The purpose of our methodological guide is to make our D-ILA data model widely known and applicable. Furthermore, it is designed in such a way that you can follow the steps we have developed while configuring the options according to your needs.

In the following chapters, we will review the basic concepts that are absolutely necessary to understand the project, then present the D-ILA data model that has been created and share the experiences we have gained from testing our D-ILA data model with AI. Finally, we discuss the policy implications of the D-ILA data model and its applicability in practice.

Please note that all results presented in the methodological guide are public and can be used freely. Please visit the official project website at www.kifu.gov.hu/d-ila. We would like to support the use of the D-ILA data model even after the end of the project, so please do not hesitate to contact us via the email address provided on the website or via the project's LinkedIn contact.

This methodological guide is a product of the project "Digital Individual Learning Accounts in the Visegrad Countries". The project is funded by the European Union (EU) and aims to support Member States in developing an enabling framework for Individual Learning Accounts (ILAs), thereby contributing to increasing the number of adult learners. The main objective of the project is to model and test a data model that would allow the use of AI technologies for personalised training recommendations and efficient spending through a delivery mode such as ILAs. Due to the type of the project (ERASMUS+ KA2, adult education), we dealt with transversal skills only.

II. Conceptual framework of the D-ILA data model

In this chapter, we briefly review the basic concepts that are necessary to understand the methodological guide. Due to space limitations, we will not go into detail on each topic. If you are interested, you can find more detailed information in the WP2 Feasibility Study[1] document prepared in the framework of our D-ILA project.

II.1 Policy modelling

In education policy, modelling can be interpreted and applied in many different ways. In general, modelling includes tools and methods used to analyse and design education systems, processes, and policies. Here are some of the main areas where modelling can be applied:

Statistical modelling and data analysis:

- Predicting student performance: Statistical models can be used to predict student performance, completion rates, retention, and other key indicators.
- Analysing educational inequalities: Data analysis and statistical modelling can be used to identify existing inequalities in the education system, such as differences based on socio-economic background.

[1] The document is available: https://kifu.gov.hu/publications/

Funded by the European Union

Systems theory modelling:

- Simulation of the education system: Modelling and simulating the different elements of the education system (e.g. schools, teachers, students) and their interactions helps education policy makers to understand how the system works and the impact of certain changes.
- Predicting the impact of policies: Simulation models can be used to predict the impact of different political measures, such as new curriculum reforms or funding changes.

Economic modelling:

- Cost-Benefit Analysis: Analysis of the economic impacts of educational investments, including long-term economic growth and labour market impacts.
- Funding models: Analysis and planning of education system funding structures, such as a combination of public grants and private sources.

Pedagogical modelling:

- Modelling learning processes: Modelling learning processes and methods helps to understand how different pedagogical approaches affect student performance.
- Curriculum Development: Analysing and modelling the effects of curricula and educational programmes in order to develop more effective educational materials and methods.

Network Analysis:

- Networks of Schools and Communities: Network analysis helps us understand how schools, teachers, students, and communities are connected and how these connections affect educational outcomes.

Modelling is therefore a key tool for education policy makers, because it enables a better understanding of complex systems and processes and helps to make informed decisions to improve the education system.

Within the framework of the D-ILA project, we dealt with predicting the impact of political measures and modelling financing among the listed ones. We formulated various policy ideas that were realistic from the perspective of the participating countries, and then translated the policy ideas into the language of our D-ILA data model. This means that the input data tables of the model were transformed as it followed from the political vision. This was followed by the running of the AI that evaluates the input data tables, and then the comparative analysis of the changes appearing in the resulting output data tables. In this methodological guide, we present in detail the modelled professional political ideas, the data tables changed for their effect, and finally the changes created in the results of the modelling.

II.2 Data in the D-ILA data model

Data often reflects a slice of reality, but it can never fully capture its complexity. The reliability and accuracy of the data largely depends on the methods of data collection. Poorly designed data collection processes or biased questionnaires can result in inaccurate or misleading data. It is therefore crucial that data collection methods are well thought out and systematic. Data plays a fundamental role in decision-making, especially in education policy. Decisions made based on data are usually more informed than decisions based on gut feeling alone. At the same time, it is important for decision-makers to look at the data with a critical eye and not to ignore qualitative information either.

Funded by
the European Union

Based on what has been described, what data we use in the D-ILA data model is particularly important. In the early phase of the project – as early as during the preparation of the application documentation - we were faced with the fact that no comprehensive and reliable data was available for the analysis of the Individual Learning Account-based financing of adult education. This is due to the novelty of ILAs, as well as the strict regulation of personal data management and, consequently, the shortcomings of national data systems. On the other hand, we believe that there will be high-quality data available if we can present the possibilities inherent in the use of data and the advantages of modelling. If the decision-makers become aware of the potential of data-based decision-making, they will strive to collect the necessary data and ensure its usability.

In order to resolve the anomaly, we worked with realistic but not real data during the D-ILA project. The data tables included in the D-ILA data model were edited by an international team of experts who, using their extensive professional experience, were able to produce realistic data. This means that when we presented the edited data to independent experts, they thought of real people and real adult training based on the data.

An important consequence of using realistic but not real data is that the obtained results are not suitable for supporting real policy decisions. That is why, during the project, we did not focus on the analysis of the data, but we recorded important moments from a methodological point of view - among other things in this methodological guide.

With this solution, we enabled entrepreneurs to use the project results to get to know the countless possibilities inherent in modelling. By filling the D-ILA data model with real data, getting real results that is, progress towards the implementation of data-based decision-making can be made.

II.3 Individual Learning Accounts (ILAs)

A conceptual framework for Individual Learning Accounts (ILAs) in Europe has been evolving over recent years, reflecting broader changes in lifelong learning policies and strategies across the EU. The idea was already being tested in many countries at the turn of the millennium. The terminology used for this instrument has been quite varied (training accounts, education accounts, training vouchers, education savings accounts, etc.), differing in a number of other specific elements in each case. Some authors strictly separate de facto accounts, i.e. an instrument for the long-term use and savings of the participant, from de facto vouchers, which give the right to use a credit (a sum of money) on a one-off basis. Regardless of the terminology, however, the basic principle is the same. A certain group of adults (or all of them) are given a sum of money (one-off, recurrent, unconditional or subject to some predefined conditions, sometimes using hours as units) which can be used mainly for education or for other related purposes (e.g. payment of education-related costs, counselling).

The idea of ILAs is based on two basic theses:

(1) Education should not end at the end of initial education, because acquired knowledge and skills become obsolete. The development of society and the world of work confronts each individual with new challenges for which previously acquired competences are not sufficient.

(2) Those who need it the most are the least educated. Virtually all data sources point to a simple fact: participation in further education is higher the higher the educational attainment of an individual. This poses a number of problems for society, because low educational attainment and low levels of competence are highly correlated with vulnerability, particularly in the labour market, but also in other areas of life.

Funded by
the European Union

In recent years, the concept of ILAs has gained considerable traction as part of the broader EU policies aimed at promoting lifelong learning and skill development. The European Commission has endorsed ILAs as a financial instrument to facilitate the right to lifelong learning and to enable individuals to engage in training that is aligned with the demands of the labour market. Consequently, the Commission has urged Member States to implement ILAs as a means of enhancing participation and motivation for lifelong learning, thereby reducing skills gaps. In a broader context, ILAs represent a policy response to the successful green and digital transitions of individuals and societies, which are underpinned by the key strategic initiatives of the EU in the areas of education, training, and employment.

The EU targets for 2030 presented in the European Pillar of Social Rights Action Plan[2] include (among others) a plan to engage 60% of adults in the EU in training each year in order to obtain relevant skills for the labour market, society, and their personal life. The concept of ILAs is thus part of a broader approach to support adaptation to the rapidly changing job market, technological advancements and societal changes that require workers and citizens to constantly update their skills.

ILAs typically involve a government-designed fund, which individuals can use to cover or co-finance expenses related to education and training. ILAs provide financial resources specifically allocated for educational and training purposes, empowering individuals to invest in their personal and professional growth. The core idea is to give individuals direct control over their educational investments, fostering a sense of responsibility and ownership over their career development[3].

OECD characterises ILAs as "virtual, individual accounts in which training rights are accumulated over time". The EU Recommendation on ILAs states that an "Individual Learning Account is a personal account that allows individuals to accumulate and preserve their entitlements over time, in order to use them for whichever eligible training, guidance or validation opportunity they deem most useful and whenever they want to, in line with national rules".

ILAs complement other funding schemes for adult learning, such as training vouchers, individual saving accounts for training, training funds, and other.

The main features and goals of ILAS are:

1. To support autonomous decisions of individual learners regarding their training activity (as well as its provider), in accordance with the regulations set forth by the state. ILAs are designed to foster individual responsibility for an individual's own learning and career planning, and employability, as well as ownership and motivation for learning regardless of recent employment status. By its nature it helps to cultivate the culture of lifelong learning.
2. To promote quality learning opportunities by ensuring that the training programmes funded through ILAs meet quality standards. The ILAs are typically accessible through a single portal, which provides information on the training programmes and providers; the quality of both is assured by a quality assurance mechanism supervised by the state.
3. To guarantee accessibility and inclusiveness for all adult learners. In many ILA schemes, additional support (both financial and non-financial) is provided to the learners facing barriers in accessing learning.

[2] https://ec.europa.eu/social/main.jsp?catId=1607&langId=en
[3] Ľ. Habodászová. 2022. INDIVIDUAL LEARNING SCHEMES – RECOMMENDATIONS FOR IMPLEMENTATION IN SLOVAKIA available at: https://doi.org/10.31577/PPFAR.2022.14.007

Funded by
the European Union

The implementation of ILAs has been observed in a number of countries, including the United Kingdom, Scotland, Austria, the United States of America, Singapore, Canada, and others, with a multitude of national elements and differences. According to the OECD “only one real Individual Learning Account exists: the Compte Personnel de Formation in France”[4] (French Personal Training Account).

In the countries of the Visegrad region, ILAs are in various phases of implementation.

In the Czech Republic, the Ministry of Labour and Social Affairs of the Czech Republic has been delivering financial support to individuals through the database of re-skilling courses and courses of further learning subsidised by the state. The scheme is available through the dedicated web-site[5] of the ministry and supported by the Labour Office branches. Only courses on digital skills and skills related to Industry 4.0 developments are subject to this support. The scheme serves as a pilot for eventual wider implementation of ILAs.

The Hungarian Government is highly committed to introducing ILAs and creating the conditions for its implementation. For that reason, the Technical Instrument Scheme of the European Commission was used: it prepares the ILAs environment and its launch. The development of ILAs is underway and the necessary regulatory environment is planned to be in place in 2024.

In Poland, the debate on ILAs has been ongoing since 2022. In 2023, a Working Group on the development of skills relevant to the labour market was established by the Ministry of Family and Social Policy. The ILA pilot is to be implemented within the framework of the European Social Fund Plus for the period 2021-2027. The final model for financing ILA schemes and attracting participation in ILAs in Poland is still being developed.[6]

In Slovakia, the Ministry of Education, Research, Development and Youth has been implementing the ESF+-funded project EPIVU – The Electronic platform for ILAs. The project will create a portal for the ILAs provision with an integrated register of quality-assured providers of the training courses. In parallel, preparation of a new legislation – the Act on Adult Learning, which will provide a legal framework for ILA implementation, is in its final stage (expected in the 2nd half 2024).

II.4 Soft skills and related frameworks

Soft skills play a pivotal role in adult education and the labour market as they are personal attributes that enhance an individual's interactions and job performance. Our project is dedicated to the development of soft skills within adult education, guided by UNESCO's definitions and expanded by other relevant skill sets, including green skills, language skills, and ICT (digital skills).

The concept of skills in general encompasses various elements of action and activity performed without direct mental control, resulting from extensive practice. According to the definition in the European Parliament's Decision 2018/646; skills cover a broad spectrum of what a person knows, understands, and can do, including knowledge, competences, and the application of know-how to solve problems and complete tasks.

[4] OECD (2019a). Individual Learning Accounts: Panacea or Pandora's Box?, OECD Publishing, Paris. Available at: https://doi.org/10.1787/203b21a8-en.
[5] https://www.mpsv.cz/web/cz/jsem-v-kurzu
[6] https://year-of-skills.europa.eu/news/individual-learning-accounts-where-are-we-now-2023-11-21_en

Funded by
the European Union

Soft skills include a wide range of abilities that go beyond technical expertise, focusing on personal attributes and social competences. Unlike hard skills, which pertain to specific technical abilities, soft skills are intrapersonal, interpersonal, and usually broadly applicable. Examples of soft skills include empathy, leadership, integrity, time management, and decision-making. These skills are often more challenging to measure and develop compared to hard skills, which can usually be demonstrated through tangible qualifications.

Although hard and soft skills are often seen as dichotomous, in reality they exist on a continuum. Effective job performance requires a combination of both, with soft skills being crucial for interpersonal interactions and hard skills providing technical expertise. Employers value the right blend of these skills, with soft skills frequently deemed more critical for success within an organisation.

Soft skills and transferable skills are often viewed as synonymous in various literature, with the assumption being that hard skills lack transferability. In reality, both soft and hard skills can be transferred across different companies, sectors, and countries. However, the degree of transferability differs. Soft skills generally have a higher level of transferability, making them essential for adaptability in the labour market. As defined by the UNESCO International Bureau of Education, soft skills encompass intangible personal qualities, traits, habits, and attitudes applicable to many job types. Furthermore, the European Commission's publication on the Transferability of Skills across Economic Sectors underscores that soft skills are non-job specific and essential for effective workplace performance, highlighting their perfect transferability. Conversely, hard skills are more technical and specific to particular occupations, resulting in a lower level of transferability. However, the concept of transferability is sometimes questioned because individuals learn to perform tasks in specific contexts, which may not always translate to different environments.

Soft skills are vital for maintaining and enhancing employability. They support specific technical skills and often make a significant difference in job retention and career advancement. Employers and employees alike recognize the value of investing in soft skills to increase adaptability and job performance across different roles and industries.

In the EU structures, several frameworks and databases have been established to include soft skills. The frameworks represent tools systematically listing and describing soft and/or transversal skills related to a specific area, enabling common referencing and levelling. The main frameworks, including DigComp, CEFR, LifeComp, GreenComp, EntreComp and FinComp, are discussed closely in Chapter II.4.2. The European Skills, Competences, Qualifications and Occupations (ESCO)[7] database serves as a comprehensive classification system that organises skills, competences, qualifications, and occupations across the EU. Within this database, transversal skills and competences are given special attention due to their broad applicability and importance in the labour market. It also encompasses digital and green skills. The DigComp 2.2 framework outlines 21 competences across five areas, integrated into ESCO's skills pillar. These competences are structured hierarchically and are accompanied by metadata, such as preferred terms in multiple languages, making them accessible and relevant across the EU.

Green skills, reflecting the EU's sustainability goals, include 571 ESCO skills and knowledge concepts. The comprehensive listing and classification of these green skills are available on the ESCO portal, and

[7] https://esco.ec.europa.eu/en/classification/skill_main#overlayspin

Funded by
the European Union

users can access detailed descriptions and metadata. This enables individuals and organisations to align their training and development programmes with sustainability objectives.

ESCO's structured classification system also facilitates the use of machine learning for mapping and matching skills and qualifications. This can be particularly useful for automating the alignment of training programmes and job requirements with ESCO's standardised vocabulary.

II.5 Soft skills frameworks in the EU

The chapter focuses on an overview of EU soft skill-relevant frameworks.

DigComp 2.2

The Digital Competence Framework for Citizens, also known as DigComp, provides a common language to identify and describe the key areas of digital competence. It is an EU-wide tool to improve citizens' digital competence, help policy-makers formulate policies that support digital competence building, and plan education and training initiatives to improve the digital competence of specific target groups.

This chapter presents version 2.2 of the Digital Competence Framework for Citizens.[8]

Structure of the framework

DigComp 2.2 defines 5 competence areas with 21 competence elements. Each competence element has 1 to 8 proficiency levels from basic to mastery.

Dimension 1 – competence areas	Short description of competence areas	Dimension 2 – competence
1 Information and data literacy	To articulate information needs, to locate and retrieve digital data, information, and content. To judge the relevance of the source and its content. To store, manage, and organise digital data, information, and content.	1.1 Browsing, searching and filtering data, information, and digital content 1.2 Evaluating data, information, and digital content 1.3 Managing data, information, and digital content
2 Communication and collaboration	To interact, communicate, and collaborate through digital technologies while being aware of cultural and generational diversity. To participate in society through public and private digital services and participatory citizenship. To manage one's digital identity and reputation.	2.1 Interacting through digital technologies 2.2 Sharing through digital technologies 2.3 Engaging in citizenship through digital technologies 2.4 Collaborating through digital technologies 2.5 Netiquette 2.6 Managing digital identity

[8] https://publications.jrc.ec.europa.eu/repository/handle/JRC128415

Funded by the European Union

3 Digital content creation	To create and edit digital content. To improve and integrate information and content into an existing body of knowledge while understanding how copyright and licences are to be applied. To know how to give understandable instructions for a computer system.	3.1 Developing digital content 3.2 Integrating and re-elaborating digital content 3.3 Copyright and licences 3.4 Programming
4 Safety	To protect devices, content, personal data, and privacy in digital environments. To protect physical and psychological health, and to be aware of digital technologies for social well-being and social inclusion. To be aware of the environmental impact of digital technologies and their use.	4.1 Protecting devices 4.2 Protecting personal data and privacy 4.3 Protecting health and well-being 4.4 Protecting the environment
5 Problem solving	To identify needs and problems, and to resolve conceptual problems and problem situations in digital environments. To use digital tools to innovate processes and products. To keep up to date with the digital evolution.	5.1 Solving technical problems 5.2 Identifying needs and technological responses 5.3 Creatively using digital technologies 5.4 Identifying digital competence gaps

Usability

DigComp is used by Europass CV Online, a self-assessment tool based on the Digital Skills and Jobs Platform, a self-checker for DigCompSat and integrated into the Digital Skills Index.
In addition, DigComp is a conceptual, reference-level framework that EU countries can use as a basis for developing their own digital competences frameworks for citizens, considering local needs.
Measuring tools: several self-assessments and a learning outcomes-based development system (the French PIX) have been completed. A technical concept for the Hungarian DigKomp Learning Support Platform has been developed, which also implements learning outcomes-based assessment.
Self-assessment tools: DigCompSat, Europass, Digital Skills and Jobs Platform, IKANOS.

CEFR

The Common European Framework of Reference of Languages (CEFR) was developed by the Council of Europe and formally published in 2001. Its main objective was to create educational and cultural convergence between the Member States in the field of foreign language teaching and to promote transparency and consistency in the learning and teaching of modern languages throughout Europe. [9]
The CEFR was first published in English and French and has since been translated into 33 languages.

[9] Source: https://www.europarl.europa.eu/RegData/etudes/etudes//2013/495871/IPOL-CULT_ET(2013)495871(SUM01)_HU.pdf [4. o.]

Funded by the European Union

Structure of the framework[10]

The CEFR defines a total of six language proficiency levels, divided into three groups. At each level, action-oriented descriptions are used to define what language learners at each level know and are able to do. These levels range from the basic knowledge of the beginner language learner to the advanced language learner.

Table: Global scale

User category	*Level*
Proficient User	C2
	C1
Independent User	B2
	B1
Basic User	A2
	A1

In addition to the six levels of criteria shown in the table, the CEFR also distinguishes three additional levels: A2+ (between A2 and B1), B1+ (between B1 and B2), and B2+ (between B2 and C1).

Table: Self-assessment grid

Competence area	*Competences*
Reception	Listening
	Reading
Interaction	Spoken Interaction
	Written Interaction
Production	Spoken Production
	Written Production

All 6 competences can be assigned a level from A1 to C2.

The CEFR alone cannot be used to measure language proficiency; it is a framework in itself and is not suitable for measurement, but a range of measurement tools compatible with the framework has been developed and no major language examination centre today will issue a language examination without indicating the level of the framework. A Level Matching Manual has also been developed to help different exam developers align the levels of their language tests with the levels of the CEFR. As a result, most language exam centres that issue language exams, language training centres advertising

[10] https://rm.coe.int/168045b15e

their courses, and even language books indicate the relevant CEFR levels. As a result, most language learners are aware of the levels' meaning.
The CEFR is not language-specific, and therefore does not contain specific vocabulary lists or grammatical structure lists broken down into levels for any living foreign language.[11]
As far as the implications regarding the development of the data model is concerned, the areas should definitely be included, both because of the training as well as the needs of employers, because some areas are communication, some are literacy, some are both.

LifeComp[12]

LifeComp: The European framework for the personal, social, and learning to learn key competence is a framework to establish a shared understanding on the "personal, social and learning to learn" key competence. LifeComp is a non-prescriptive conceptual framework that can be used as a basis for the development of curricula and learning activities. The aim is to build a meaningful life, cope with complexity, be thriving individuals, responsible social agents, and reflective lifelong learners. LifeComp describes nine competences that can be learned by everyone in formal, informal, and non-formal education.

The framework describes nine competences (P1-3, S1-3, L1-3), which are structured around 3 interlinked areas of competence:

Table: The structure of LifeComp

Competence area	*Competences*
Personal	P1 self-regulation
	P2 flexibility
	P3 wellbeing
Social	S1 empathy
	S2 communication
	S3 collaboration
Learning to learn	L1 growth mindset
	L2 critical thinking
	L3 managing learning

Instruments to measure personal, social, and learning competences can be used to determine whether an individual has them. More sophisticated use would require levels and level descriptions, which are

[11] http://www.keronline.hu/ There is no official list, but various publishers have already published a large number of lists giving suggested vocabulary sets for different levels. This is more positive than if something had been published centrally by the creators of the CEFR, because it demonstrates the acceptance (and use) of the framework. They are also the basis for the EUROPASS CV and are used elsewhere to identify or assess an individual when they are recruited for a traineeship. For employers this is an important factor, it should not be left out of the data model in our opinion. See here, too: https://nyelviskola.hu/kozos-europai-referenciakeret-szintek

[12] https://joint-research-centre.ec.europa.eu/lifecomp_en

Funded by
the European Union

not currently available from the information available. To our knowledge, no self-assessment or measurement tool has yet been developed for the framework. The next step in the development of LifeComp is to field test the framework, implement it in a real environment and evaluate the results.[13]

GreenComp[14]

GreenComp defines sustainability competences that can be integrated into educational programmes to help learners develop knowledge, skills, and attitudes that promote thinking, planning, and action with empathy, responsibility, and care for our planet and public health.

Structure of the framework

"GreenComp comprises four interrelated competence areas: 'embodying sustainability values', 'embracing complexity in sustainability', 'envisioning sustainable futures' and 'acting for sustainability'. Each area comprises three competences that are interlinked and equally important."[15] GreenComp consists of 12 competences organised, into the four areas below:

Table: The structure of GreenComp

Competence area	*Competences*
Embodying sustainability values	valuing sustainability
	supporting fairness
	promoting nature
Embracing complexity in sustainability	systems thinking
	critical thinking
	problem framing
Envisioning sustainable futures	futures literacy
	adaptability
	exploratory thinking
Acting for sustainability	political agency
	collective action
	individual initiative

[13] Source: Sala, A., Punie, Y., Garkov, V. and Cabrera Giraldez, M., LifeComp: The European Framework for Personal, Social and Learning to Learn Key Competence, EUR 30246 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19418-7, doi:10.2760/302967, JRC120911. 16. p.

[14] https://joint-research-centre.ec.europa.eu/greencomp-european-sustainability-competence-framework_en

[15] Source: https://publications.jrc.ec.europa.eu/repository/handle/JRC128040

Funded by
the European Union

Custom-developed measurement tools specific to the competences of the European Framework of Competence for Sustainable Development can be used to assess whether or not an individual has the competences for which no level definitions are available. No measurement tool has been developed for the framework. No self-assessment tool has been developed for the framework.

“Although widely endorsed by subject-matter experts and representatives of different stakeholder groups, the framework has not yet been tested in a real setting. Putting GreenComp into practice, by rolling it out and evaluating it in a specific context, could and should lead to amending and refining it based on feedback from practitioners and end users. The framework should thus be treated as a living document.”[16]

EntreComp[17]

Developing the entrepreneurial capacity of European citizens is one of the eight key competences for lifelong learning. Entrepreneurial value creation and entrepreneurial learning can take place in any walk of life; turning ideas into shared value is important for career development. EntreComp describes entrepreneurship as a lifelong competence and identifies the elements that make someone an entrepreneur.

Structure of the framework

The EntreComp consists of 3 interrelated and interconnected competence areas: ‘Ideas and opportunities’, ‘Resources’ and ‘Into action’. Each of the areas is made up of 5 competences which, together, constitute the building blocks of entrepreneurship as a competence. The framework develops the 15 competences along an 8-level progression model and proposes a comprehensive list of 442 learning outcomes.”[18]

Table: The structure of GreenComp

Competence area	*Competences*
Ideas and opportunities	spotting opportunities
	creativity
	Vision
	valuing ideas
	ethical and sustainable thinking
Resources	self-awareness and self-efficacy
	motivation and perseverance

[16] Source: Bianchi, G., Pisiotis, U., Cabrera Giraldez, M. GreenComp – The European sustainability competence framework. Bacigalupo, M., Punie, Y. (editors), EUR 30955 EN, Publications Office of the European Union, Luxembourg, 2022; ISBN 978-92-76-46485-3, doi:10.2760/13286, JRC128040. 9. p.

[17] https://entrecompeurope.eu/wp-content/uploads/EntreComp-A-Practical-Guide-English.pdf

[18] Source: Bacigalupo, M., Kampylis, P., Punie, Y., Van den Brande, G. (2016). EntreComp: The Entrepreneurship Competence Framework. Luxembourg: Publication Office of the European Union; EUR 27939 EN; doi:10.2791/593884. 0. p. Abstract

Funded by
the European Union

	mobilising resources
	financial and economic literacy
	mobilising others
Into action	taking the initiative
	planning and management
	coping with ambiguity, uncertainty, and risk
	working with others
	learning through experience

EntreComp contains definitions of competences and associated level definitions. For each competence, the same levels are defined. They are numbered 1 to 8 and define proficiency levels: basic (1 and 2), intermediate (3 and 4), advanced (5 and 6) and expert (7 and 8). The definitions can support the development of measurement tools. The competences of the Entrepreneurship Competence Framework can be measured using custom-developed measurement tools. No self-assessment tool has been developed for the framework. A subsequent step will be to try the EntreComp Framework out in practice, by implementing and evaluating it in a specific context and, eventually, amending and refining it according to feedback from practitioners and end-users if necessary.[19]

FinComp[20]

The meaning of FinComp abbreviation is: Financial competence framework for adults in the European Union. "The objective of the EU/OECD-INFE financial competence framework for adults is to promote a shared understanding of financial competences for adults amongst Member States and national authorities, educational institutions, industry, and individuals. In addition, it provides a basis for a more coordinated approach among EU and national policymakers. By supporting efforts to improve financial literacy, the framework aims at contributing to the overall goal of improving individual financial well-being."[21]

Structure of the framework

FinComp divides the competences into four content areas: money and transactions, planning and managing finances, risks and reward, and financial landscape[22]. These content areas have then been further divided into topics and subtopics.

[19] Source: Bacigalupo, M., Kampylis, P., Punie, Y., Van den Brande, G. (2016). EntreComp: The Entrepreneurship Competence Framework. Luxembourg: Publication Office of the European Union; EUR 27939 EN; doi:10.2791/593884. 9. p

[20] https://www.oecd.org/finance/financial-competence-framework-for-adults-in-the-european-union.htm

[21] Source: European Union/OECD (2022), Financial competence framework for adults in the European Union. 2. p.

[22] This structure is in line with the previous OECD/INFE Competence Framework for Adults.

Funded by
the European Union

Table: The structure of FinComp

Content area	*Topic*
1. Money and Transactions	1.1 Money and Currencies
	1.2 Income
	1.3 Prices, Purchases, and Payments
	1.4 Financial Records and Contracts
2. Planning and Managing Finances	2.1 Budgeting
	2.2 Managing Income and Expenditure
	2.3 Saving
	2.4 Investing
	2.5 Longer-Term Planning and Asset Building
	2.6 Retirement
	2.7 Credit
	2.8 Debt and Debt Management
3. Risk and Reward	3.1 Identifying Risks
	3.2 Financial Safety Nets and Insurance
	3.3 Balancing Risk and Reward
4. Financial Landscape	4.1 Regulation and Consumer Protection
	4.2 Rights and Responsibilities
	4.3 Financial Education, Information, and Advice
	4.4 Financial Products and Services
	4.5 Scams and Fraud
	4.6 Tax and Public Spending 4.7 External Influences

Unlike the frameworks presented earlier, FinComp has hundreds of competences, which makes its use in ILAs questionable. At the same time, each indicator is assigned a cross-cutting dimension, of which there are only 5. As a result, the dimension consisting of 5 elements can now be used well in the D-ILA data model to characterize competences.

Funded by
the European Union

Table: Tags in the FinComp framework

Digital financial competence
Sustainable finance competence
Competence relevant to financial resilience
Competence relevant for daily life competence and/or for current or future financial well-being
Competence relevant to a large majority of the adult population

Several labels can be assigned to a FinComp competence. Each label can have a value of 0 or 1. For the most general competence, the value of 5 labels can be up to 5 values of 1.

This framework is a tool to support policy makers and practitioners in the creation of their own policies and programmes, rather than a curriculum, but it can easily be adapted to address the needs of specific life situations or target groups. For instance, future users of the framework will be able to select and extract the most relevant competences for some specific "life stages". Another possible use of the framework can be to select and extract the most relevant competences for some specific target groups, such as women, seniors, young people, low-income groups or other groups who may be financially vulnerable.[23]

No measuring tool was made for the framework. No self-assessment tool has been developed for the framework.

II.6 Application of Artificial Intelligence in Modelling

Artificial Intelligence (AI) is a general-purpose technology that can be used for the automation of processes, data analytics, and supporting human decision-making. AI refers to the field of computer science and technology that aims to create intelligent machines capable of simulating human cognitive functions. The umbrella term AI may refer to various technologies, such as machine learning, deep learning, neural networks, and natural language processing.

Modelling of data is understood as creating mathematical or computational representations of real-world processes and systems based on collected data. This practice aims to capture the essential characteristics and relationships within the data to understand, predict, and simulate behaviours and outcomes. Data modelling can take various forms, including statistical models, machine learning algorithms, and simulation models. These models help in uncovering patterns, testing hypotheses, and making informed decisions by providing a structured framework to analyse complex datasets, thereby enabling a deeper understanding of the underlying phenomena.

AI is extensively utilised in the modelling of data and decision-making processes, because it can analyse large datasets to identify patterns, make predictions, and generate insights that inform decision-making. By learning from historical data, AI models can help in forecasting future trends and optimising strategies, as well as automating complex decisions quickly and with sufficient accuracy.

[23] Source: European Union/OECD (2022), Financial competence framework for adults in the European Union 6-7. p.

Funded by the European Union

In the D-ILA project we envisaged modelling as one of the main outputs. On the one hand, this is a way of showing the potential of digital ILAs, on the other hand, machine readability supports policy modelling, thereby making better informed policy decisions.

In applying the AI to modelling, we applied the following logic. First, we gathered sufficient data on training/learning opportunities, learners' characteristics and their choices. We structured the data in our data model to describe key variables such as cost/availability of funds, time availability, course duration, and many others. All of which are strictly related to the use of ILAs and adult learning. Using this collected data, we then train an AI to find the hidden patterns matching learners' characteristics with the training attributes. We use the AI to model the outcomes of matching when specific variables are changed, for example: the amount of entitlement for individuals in the ILAs, the quantity of training offers, the availability of time for learning etc, or when a new learner or training is entered. Finally, these results can be used in two ways. For individual cases (for example, becoming a training recommender for individuals) and on a policy level - to assess possible results of various choices on the whole population.

Within the project, synthetic data on learners and training offers has been created by experts in the field of education and adult learning. The algorithm used has been trained on this data (300 individuals' data and data describing 100 courses) and will generate results, which have the limitations of the synthetic data. However, the model allows using training course data, so whenever higher quality input data is available, the results should increase in quality as well.

III. Preparations for the creation of the D-ILA data model

In the initial phase, the project evaluated how an ILA system could be implemented in the Visegrad region (Czechia, Hungary, Poland, Slovakia) by examining the current state of adult education and training, particularly in terms of soft skills and the processes determining matching the right individuals with the right training. The potential uses of AI in matching training programmes to learners' profiles were also evaluated.

III.1 Conclusions on adult learning systems and their funding solutions

The conclusions that can be drawn from the analysis of adult education funding systems, including the situation of ILAs, are as follows:

- Based on the analysis of adult education systems, it can be concluded that each country is building its own systems according to very different philosophies. It follows that, because of these different philosophies, the practical implementations also differ.
- This discrepancy is also reflected in the availability of data. From an ILA data model perspective, it is currently difficult to develop a common dataset that can be used in various countries.
- The ILA data model may reflect a future state that countries aspire to achieve.
- In our view, the EU concept of ILAs is general enough to be integrated into the adult education systems of all the countries analysed.

Funded by
the European Union

- Adaptation to local characteristics implies that interoperability of ILAs between countries can only be achieved with great difficulty, considering the conclusions drawn from the feasibility study.
- The question for us is whether the EU concept of ILAs (as well as micro-credentials) and the EU resources allocated to its implementation are capable of influencing the further development of each country's own adult education system.
- A systemic difficulty in implementing ILAs is that EU projects tend to be one-off initiatives that lack the necessary continuity.
- Enterprises use their own training solutions. On the one hand, this is advantageous due to the applicability of the ILA data model, as data requirements are significantly easier to implement than in large public systems. On the other hand, it is disadvantageous from an ILA point of view: enterprises pay for training on their own initiative and have no interest in transferring money to employees' learning accounts. Whether employers participate in the ILA system depends on the country that is implementing the ILAs; there are countries where not only the state, but also the employer can contribute to an individual's learning account.
- There is a need to increase the participation in adult education in the four countries studied. One obstacle to this is the lack of financial resources, especially for disadvantaged people.
- However, we believe that the reorganisation of financial resources is not in itself a solution to increasing the participation in adult education.
- In addition, given the current specificities of adult education systems, we are not convinced that individuals will be able to effectively manage the budget allocated to them from different sources.
- We believe that it is very important that widespread support services are put in place in parallel with the implementation of ILAs. In the countries studied, we do not believe that the support services needed for ILAs are currently in place.
- Our analysis shows that the inclusion of micro-credentials in the ILA data model is important as adult education systems in the countries studied have moved towards shorter courses. Digital micro-credentials are primarily suitable for certifying the learning outcomes of micro-training.
- The situation is further complicated by the fact that international corporations, as market players, also issue micro-credentials to promote their technology and develop potential employees in response to emerging needs.

III.2 Conclusions based on the presentation of national soft skills solutions

- In general, the management of soft skills training has not yet been established in the countries studied, but the first steps have already been taken.
- Soft skills are managed using existing and older systems, with an expansion of its functions. First of all, adding soft skills to the existing characteristics of the occupations.
- Soft skills management solutions primarily reflect labour market needs, which place increasing demands on the adult education system.
- A general aim is to assign levels to competences and ensure their measurability.
- At the national level, the most accepted approach to soft skills emphasises transferability between jobs and occupations, a transversal approach.

Funded by
the European Union

- An important starting point in the countries studied is the concept of lifelong learning, which is also accompanied by the development of soft skills.

III.3 General findings on the ILA data model

Our general findings on the ILA data model are summarised in this chapter based on the background study prepared by the Partnership within the framework of the project:

- The adult education system and related data system of each country largely determines the availability and structure of data relevant to the ILA data model. These vary considerably from country to country. Therefore, the EU Recommendation on ILAs has been given high priority in the development of the data field structure of our ILA data model. Frameworks defining levels of competences in five areas (digital, green, entrepreneurship, life, and language) were used as a guide to identify the level of skills to which the educational offer is intended to contribute.
- The differences between national adult education systems have implications for the availability and structure of data relevant to the construction of an ILA data model. As a consequence, the proposed data model will be aligned with EU recommendations and international frameworks. This means that, while the proposed model may include a controlled vocabulary of skills based on EU competence frameworks such as DigComp or EntreComp, we expect that future implementations of ILAs in Member States will use national classifications of skills, occupations, qualifications, and national competence frameworks where these exist. This decision assumes that Member States will implement or refer to these reference frameworks (as in the case of NQFs-EQF).
- The background studies carried out as part of the project revealed that state and corporate spending on soft skills development is directed at significantly different target groups.
- Corporate resources serve the upskilling and/or retraining of employees, while public spending has a broader scope, e.g. including social inclusion. Although the ILAs as a delivery mode foreseen in the Recommendation are independent of the funding source and provide full control to the individual, different aspects can be emphasised in various datasets. Moreover, Member States may decide to develop a variant of ILAs in which the individuals' spending decisions will not be completely independent of the funding or individuals will be provided with recommendations based on a specific perspective. The key issue to be decided during the development of the ILA data model is whether to construct the participants' data series of the AI training data table and the data series of the test data table exclusively from a corporate or state perspective, or from a mixed perspective. We did not find such a sharp difference in the case of training courses, due to the transversal nature of soft skills training, where a specific type of training may be relevant from both corporate and state perspectives.
- At the corporate level, especially Small and Medium Enterprises (SMEs), there is a better chance of creating an abundance of data – if the return on investment (ROI) and corporate interests can be demonstrated (employees can easily give their consent to the management of their data).
- Previously, professional competences related to specific professions were considered significantly more important by enterprises than soft skills training. Based on the available data, changes can be observed: in the context of developments in the labour market, society,

Funded by
the European Union

and technological advancement, enterprises are placing increasing emphasis on soft skills development. Moreover, in many cases, recruitment has become soft skills based rather than focused on existing professional knowledge. Our ILA data model deals exclusively with the development of soft skills.

- The above-mentioned development is not so much related to EU efforts as to changes in the labour market. It is difficult for enterprises to find an individual with the right hard skills, either because of (a) the general lack of qualified workers available (due to low unemployment), or (b) the nature of the technological developments where the requirements for technical skills are so specific (often even company specific) that it is not possible to find the right person available in the labour market. At the same time, many jobs require more soft skills due to the same technological developments (everything is more connected, collaboration and willingness to constantly learn is needed etc....). For all these reasons, enterprises are placing more emphasis on soft skills than before.
- A general guideline for the development of the ILA data model is that the funder of training can influence the selection of the courses available throughout the scheme according to its priorities. For example, in the case of public funding, the same individual would receive a different training offer than if the training were funded by that individual’s employer. A rare exception is when a company funds adult education in such a way that its employees are free to choose.
- To train an AI, it is necessary to generate a data table in which individuals are assigned training courses. A very serious dilemma when designing an ILA data model is whether to model the real state or the desired one. For example, real statistics show that individuals with higher qualifications are more likely to participate in adult education. So should we also edit the data table needed to train the AI in this way? Or, conversely, should we assume that individuals with lower qualifications are more likely to need training?
- These two bullet points above show that, in short, the key decision point is the goal of the training. What type of training is recommended is (should be) determined by what we want to achieve (employability/social inclusion/company profit/good economy).
- Every person should receive the adult education they need. An important principle in the development of our ILA data model is that we consider limited financial resources. The crux of the project is the lack of a financial framework to ensure that every person receives the training that best suits them – compromises are necessary. Our project seeks to identify the effects of these trade-offs at the data level.
- Throughout the development of the ILA data model, efforts should be made to ensure that there is a one-to-one correspondence between the data fields describing the characteristics of the learners and the properties of the training, and the value sets of the data fields (the values that the data field can take). One of possible approaches is to use EU frameworks for specific competences, such as the CEFR for foreign language competences. By using the CEFR in the ILA data model, it is possible to define an individual's foreign language competence, the input expectation of a given type of training and the level of competence that can be achieved upon completion of the training, as well as the level of foreign language competence required for a job.
- The measurement of soft skills, especially their level, is a general challenge that needs to be solved in the near future. During the development of the ILA data model, we dispense with

measurement difficulties and deliberately analyse the future state when the measurability of soft skills is possible.

- Based on the CEFR mentioned above, there are very accurate and usable measurement tools.
- We expect a similar framework for other soft skills, just as we can already see the efforts of DigComp.
- GDPR requirements must also be considered when implementing ILAs. At the same time, we are structuring our ILA data model in such a way that the individuals concerned consent to the purpose-related processing of their data in the ILA data model. It is important to note that this is partly the case now, since, for example, personal, GDPR-sensitive data is also collected during the provision of training grants provided by the EU.
- An important limitation of our data model is that we used synthetic data regarding courses which individuals can apply for; therefore, when describing a course in the D-ILA data model, real-world training with partially or completely identical parameters must be searched for in the adult learning market.
- The ILA data model seems to be easily applicable in the field of career guidance, where the user is not given a specific training recommendation, but a type of training. This makes it much easier to find a relevant training course.
- In the table used to define the ILA data model, we also indicate data fields that were considered relevant during our analysis but were not included in the data model. This is done to allow users to activate the data fields listed here according to their needs during the practical implementation.
- According to the EU Recommendation, in the case of ILAs, the budget for individuals might change dynamically over time as income is received in ILAs and spent on training. In our ILA data model, we record the specific point in time when the available budget is recorded at the individual level, and from there we consider training expenditure one year after the fixed point in time. This makes it possible to assign multiple training courses to an individual.

III.4 Recommendations

The analysis of adult education systems across the Visegrad countries reveals several common challenges, particularly in terms of access, structure, and the provision of services. There is no unified, systemic approach to adult education in the participating countries, which often leads to inconsistent access to education, varying quality of training courses available, and inadequate support services.

Main Challenges in Adult Education

Lack of a Unified System: Across the Visegrad countries, the adult education landscape is fragmented, with private providers dominating certain areas. However, these private entities, especially those offering soft skills development programmes, are not always accessible to all individuals due to high costs or limited availability.

Limited Access and Awareness: Adult learners, especially in rural or underdeveloped areas, often lack access to comprehensive databases of available training courses and providers. This limitation is compounded by a lack of service and guidance for individuals seeking further education, particularly in understanding which courses are best suited to their career progression.

Quality Assurance Issues: Quality control in adult education is not consistently regulated. There are gaps in ensuring that training programmes meet specific standards, and there are no universally accepted frameworks for evaluating the quality of providers and their programmes.

Addressing Barriers to Education

Several structural barriers to participation in adult education have been identified, including geographic and socio-economic disparities. Commonly, individuals who reside in urban areas and those with higher educational qualifications are more likely to engage in lifelong learning, while those in rural regions or with limited education are often left behind.

The D-ILA model provides a potential solution to some of these issues by using AI-powered recommendations to match individuals with appropriate training programmes based on their needs and qualifications. This system has the potential to increase access to adult education by offering personalized suggestions and reducing barriers related to location and educational background.

Recommendations for Improving Adult Education Systems

The project highlights the importance of an integrated approach to adult education that includes public registries of recognised training opportunities, quality assurance mechanisms, and access to guidance and support services. Furthermore, the following key recommendations were made:

Create National Portals: Countries should develop single national digital portals that provide accessible, up-to-date information on all available training programmes, courses, and providers. These platforms should be universally accessible, including to individuals with disabilities, and should ideally be integrated with the Europass platform to ensure consistency across Europe.

Establish Quality Assurance Frameworks: A national system for quality assurance based on reliable registers of training providers should be established to ensure that individuals can trust the training they choose to pursue. Such a system should evaluate providers based on standardized criteria and make this information publicly available.

Focus on Inclusivity: Special attention should be given to non-privileged groups, including individuals living in rural areas, those with non-traditional employment contracts, and those with lower educational attainment. Programmes should be designed to cater to these groups, offering financial support and targeted outreach to ensure their participation.

Employer Involvement: Employers, especially SMEs, should be encouraged to actively participate in the development of their employees through training programmes. While larger multinational companies may have the resources to assess employee development needs, SMEs can benefit significantly from the D-ILA data model's personalised training recommendations.

Public-Private Partnerships: Funding for adult education should be a shared responsibility between public institutions, employers, and individuals. The D-ILA model offers opportunities for collaboration between these stakeholders, allowing for shared costs and enhanced access to training opportunities.

Utilising the D-ILA Data Model

The D-ILA model provides a foundation for improving the accessibility and personalisation of adult education. By integrating AI technologies, the model can offer tailored training recommendations, making it easier for individuals to identify the courses that best suit their needs. This approach has the potential to reduce the structural barriers to education and improve outcomes for adult learners.

Furthermore, the data collected through the D-ILA model can offer valuable insights into learning trends, participation rates, and the effectiveness of various training programmes. Policymakers and educators can use this data to refine the education system and ensure it meets the needs of a diverse population.

Conclusion

The project showed that approaches, stages of implementation and leading principles for ILAs in the Visegrad countries vary significantly. It confirmed that a coordinated and comprehensive approach to adult education at the national level is necessary. By leveraging AI technologies and developing national platforms for training information and quality assurance, the D-ILA project offers a pathway toward more inclusive and effective adult learning systems. The next steps should focus on implementing these recommendations, with an emphasis on inclusivity and collaboration across sectors to ensure that all individuals have access to the education they need for personal and professional growth.

III.5 Findings concerning the development of the ILA data model

This chapter summarises the main conclusions of the feasibility study. In all cases, a brief formulation of the conclusions appears in the title of the section, and they are explained and elaborated in the text of the section.

The feasibility study and professional consultations realised in the initial phases of the project led to the conclusion that the goal formulated in the project application can be implemented according to the original ideas. At the same time, it is important to note that the definition of the target group and the field of application play a significantly greater role in the design of the data model than was originally envisaged. The universality of the data model is therefore questioned on the basis of the feasibility study and lead to the recommendation that different models might be suitable for different educational goals.

Although the analysis of 42 interviews conducted in 4 countries as part of the feasibility study suggested that the resulting products can be used well in practice, an important conclusion of the feasibility study was there are limitations to the introduction of ILAs:

- Adult learning systems in the countries studied vary considerably. The differences are not only technical, but also reflect fundamental differences.
- Efforts to introduce ILA schemes also vary. Among the countries studied, there are examples of the full autonomy of learners in terms of training choices. Other ILA schemes regulate the use of funds according to national priorities, e.g. to support specific target groups or skills.
- Differences between countries are also reflected at the data level, which makes interoperability at EU level difficult, therefore, flexible metadata fields and datasets should be defined.
- Some respondents point out that the current adult education systems would benefit from ILA schemes by generating learning opportunities, as adult education organisations do not receive direct funding. Due to the lack of direct funding, adult education organisations may have a counter-interest in the full implementation of ILAs. In other countries the adult education providers do benefit from the ILA concept. They are able to attract more participants, as the participants make use of the ILA financial support.

Funded by
the European Union

- During the interviews, most representatives of multinational enterprises claimed that they have their own resources for the soft skills training of their employees, which is planned for in the annual budget. Based on this, there are methods for assessing the employees' competences and a methodology for training and matching employees. These are rather "soft" methods such as interviews and questionnaires. Systematic or data-driven schemes are very rare, even in large enterprises.

In summary, without further development of the adult learning systems in each country according to common principles, we see limited possibilities for transferring financial entitlements between countries, which is one of the potentials to be considered in the ILA concept. At the same time, based on the feasibility study, ILA solutions based on country-specific characteristics can be successfully applied as one of the tools to increase the number of participants in adult education.

Full implementation of ILAs leads to a paradigm shift

In the feasibility study, we look at the funding solutions for adult education in the countries studied. In our view, the full introduction of the concept of ILAs would lead to a paradigm shift compared to the current situation, i.e. it would fundamentally and substantially change the funding model of adult education by emphasising the individual's role as the recipient of funds for training at the expense of institutions as recipients and, indirectly, of the adult education system itself.

It is important to note that many countries are currently exploring the feasibility of the EU approach. It is far from being decided whether ILAs will be supported as an exclusive funding model or whether it will be introduced in parallel with current funding methods.

Current adult education in the countries studied is characterised by the fact that the methods and sources of funding adult education vary from country to country. A significant proportion of the training is free of charge for the learner or requires a very small contribution. In some countries, state subsidiaries are given directly to training organisations, while enterprises also finance training organisations directly. In the case of fully self-financed training, the adult education services provided by the training organisation are financed directly by the trainee from their own resources and/or training loans. Commercial training organisations providing training for direct remuneration are an integral part of education systems. There are also a number of mixed solutions, for example, a particular enterprise receives state aid for the training of its employees.

According to the EU concept of ILAs, funds for adult learning can be allocated based on the individuals' choice of training. In this case, State, employers', and private resources are concentrated in one place. It is up to citizens to take advantage of the opportunities provided by these resources. An effective guidance system is needed to influence the efficient decision-making of individuals.

It is important to note that the EU concept of ILAs does not exclude the principle that the individual or organisation funding the training should ultimately have a say in the choice of training. This is likely to be the biggest difference between countries if ILAs are fully implemented. In one case, citizens can spend the amount allocated to them through the ILA scheme in the adult education market without any restrictions. In other cases, the state/company determines the specific training course it can be spent on (of course, both solutions can be implemented with the help of an information system).

In practice, various solutions are likely to prevail. For example, there may be a limit on the amount allocated to an ILA which can only be spent on accredited courses. Another solution is for a company to cover part of the training costs for its employees and guide the choice of training for employees from a recommended pool of training courses. The State can also specify target groups for training, such as certain vulnerable groups.

Differences between the ILA data model and company training

Consider that the state or enterprises involved in funding adult education may have an influence on the target group and on the courses chosen by the target group. In this case, it is essential to create a data model that reflects the needs of the parties involved. What are these needs?

Some of the data fields may be more useful for the public sector, e.g. policy priorities are often defined in terms of labour market status, demographic, social, and educational characteristics. Other data fields may be more useful for enterprises who have a competence profile of their employees.

The logic behind the funding decisions is also different in the two cases: Adult education courses organised by the public sector and funded through ILAs may include lower qualifications to support low-skilled individuals. The financial contribution may also be limited. Based on the feasibility study, there should also be data series representing individuals where both the participant´s own contribution and the corporate subsidy amount to 0 EUR.

Therefore, the data field requirements and edited datasets of the public and corporate sectors in relation to ILAs differ both in terms of target group and soft skills training. This assertion is supported by the interviews conducted.

Based on the feasibility study and the non-profit orientation of the project, it is recommended to develop the model primarily for public systems. However, the model can also be used to a limited extent in the corporate environment.

The importance of soft skills training is growing within adult education

Based on the available statistics on adult education, there is no evidence of a breakthrough in soft skills courses and enrolments or in the individuals applying for them. At the same time, our interviewees all emphasised the importance of soft skills in the labour market and predicted that soft skills courses would continue to grow in importance in the near future. They drew attention to the fact that the development of soft skills is fundamentally different from the profession-specific approach of adult education:

- Soft skills are usually much more "elusive", i.e. they are more difficult to measure and ascribe a level to, and their use in the short or long term is not obvious. In order to increase the role of soft skills training in adult education, it is necessary to develop methods for measuring and assigning a level to and quantifying the impact of training.
- Soft skills are often mixed with personal characteristics and attitudes (or even values). This means that they are not easy to change or develop, since in some cases they require a change of personality.

Some respondents argued that due to the transversal nature of soft skills (non-discipline-related, widely applicable), the approach based on measurability and impact assessment should be abandoned and citizens should be given the opportunity to develop soft skills regardless of these aspects. They were of the opinion that a less targeted development of soft skills would have indirect social and economic effects.

Funded by
the European Union

However, based on the results of the feasibility study, the development of the ILA data model focuses on measurability and the consideration of the impact of training. The main reason for this is that ILAs are a financial solution, so it is advisable to take a financial approach. The interviews confirm this idea: the impacts (measures and quantities translated to the level of data) are to be considered before a decision is made on the amount of money available for adult education.

In the ILA data model, positive discrimination is recommended

One of the reasons for creating the ILA data model is the constraint of financial resources. If financial resources were unlimited, everyone would be able to participate in the amount and level of soft skills training that best suits them for the rest of their lives. The ILA data model and similar solutions make it possible to model the range of training that can be undertaken at the individual level, given financial constraints. Conversely, what is the total cost of training allocated to a particular target group, broken down into state, company, and individual costs?

If we consider the limited financial resources, the methodology for allocating the financial resources automatically arises, which we must also follow when designing the data series of the ILA data model.

In the background studies prepared for the feasibility study, we examined at a data level how the adult education systems of each country allocate the available financial resources. Based on the results, it appears that resources are not necessarily allocated in the most appropriate way for social and economic goals in adult education systems. Based on our analysis, the most obvious example is that disadvantaged groups are under-represented in adult education, while participation in adult education is clearly a gateway for them, i.e. it is extremely important from a social point of view. (Actually, this is not a specific statement; a multitude of analyses come to similar conclusions. The EU concept of ILAs also aims to change the current situation and increase the number of participants in adult education according to the criteria of equal opportunities).

Based on the above, it is important to note that we are not trying to map the characteristics of existing adult education systems in the ILA data model. We are working with participant data modelled along social goals. The guideline used in the development of the ILA data model can be interpreted simply as positive discrimination. An example of this is that, although the proportion of disadvantaged groups in adult education is relatively low based on real statistical data, they appear in higher numbers in the ILA data model.

Desired data connections can be established between different EU frameworks

One of the most important features of the ILA data model is that it establishes a data link between the existing competences of citizens, the input and output competences of training, and the competence expectations of society and/or enterprises. To achieve this, the soft skills in the ILA data model must be well defined, non-overlapping, and levelled.

The feasibility study will look at various solutions, of which the CEFR seems ideal for establishing a data link. The CEFR, and the assessment and development tools based on it, can help to define the target foreign language competence level (e.g. job expectations). In comparison, the current level of foreign language competence of citizens can be established. The input level of a training course can be defined, based on which the training can be selected. The learning outcomes of the training can be specified, which must be in line with the objectives set by the CEFR. This method can be used to recommend not only individual courses, but also learning pathways by building courses on top of each other.

Funded by
the European Union

The feasibility study presents other EU reference frameworks relevant to soft skills. According to our analysis, the reference frameworks are at very different stages of development, but their common feature is that soft skills are well defined and form a system. It is also important to note that the reference frameworks reflect EU ambitions.

The role of AI should be made transparent

In the feasibility study, we pointed out that advanced AI-based solutions are increasingly able to find and process information sources related to individuals and adult education, and to provide specific training to real individuals.

Previously, this required basically structured databases that were more or less understandable by humans, so the operation of the AI could be verified. The situation has now changed radically. Many AI algorithms are not transparent, it is not always clear which variables were key for a given result, such operations cannot be tracked and controlled with the involvement of human resources. There is a risk that individuals and adult education will be paired with the use of AI without human influence and transparency.

The ILA data model can be used to address the above anomaly.

The feasibility study shows that a new, previously hidden application of the ILA data model can be developed. Namely, that the application of AI can take place in 2 phases. In the first phase, the AI is responsible for filling in the ILA data model tables for individuals and training using the widest possible range of unstructured information sources. At this point, it is possible for a human to intervene, check and, if necessary, modify the competences and competence levels established by AI. This ensures transparency and enables intervention. Based on the interviews conducted, we know that the process is similar in enterprises: the manager and the employee discuss the employee's competences profile together and determine the soft skills development needs of the employee based on this. In phase 1, the AI essentially acts as a complex measurement tool.

In phase 2, the ILA data model, which has been verified by stakeholders and modified, if necessary, serves as the sole source of information for the AI. Based on the completed ILA data model, the AI matches participants and courses. This project focuses exclusively on phase 2, as the ILA data model is populated with data by Partnership experts rather than by the AI.

Of course, it is debatable whether AI has really reached the level of development outlined, but in our view, it is a fact that rapid development is heading in this direction.

IV. Introducing the D-ILA data model

The D-ILA data model provides potential users with a tool suitable for adult education planning. This allows users to create and operate their own data model based on this methodological guide. In addition to the methodological guide, other outcome products of the project will also provide support.

In this chapter, we want to help users choose the data fields they need for their model. In order to make this possible, we present in detail what considerations we followed within the framework of the project, based on what criteria and arguments we made the decisions we made. Users must take these considerations and decisions into account when building their own data model.

Funded by
the European Union

IV.1 Compilation of data fields

The following example illustrates the activities involved in compiling data fields. Similar considerations can be used to review data fields in the D-ILA data model for any organisation that seeks to apply the D-ILA data model.

In our team of experts, there was no consensus on whether the spoken language of the trainees and the language of the training should be included in the D-ILA data model. However, there was agreement that the data tables of participants and training should be treated equally. Either the language (the language spoken by the trainees and the language of the training) is included in both, or neither is included. If it had been included in only one of the data tables, it would not have been useful in assigning training to individuals.

In practice, mapping means matching data sets describing training characteristics to data sets describing trainees. In this case, the spoken language and the language of the training are a strong constraint, according to the findings of our team of experts. Strong correlation means that if one specifies the language spoken by the participants and the language of the training for the training, then a mandatory match is required during matching. Therefore, when training AI, one cannot specify any cases where there is no match between the spoken language and the language of the training. For example, it is not conceivable that a Hungarian speaker should participate in training available in Polish. Since AI learns based on the 300-row personal and 100-row training data tables provided by experts, this also prevents this from happening when running AI in the future.

In the D-ILA data model, the language field was finally included in both the data table of the trainees and the data table of the training. The main reason for this was that the use of the spoken language data field seemed very good to characterise disadvantaged people. In our view, one of the manifestations of disadvantage in the D-ILA data model is that no spoken language other than the mother tongue appears. Obviously, the reverse is also true: in the case of a highly qualified person who wants to undergo further training, we can indicate higher social status with several languages spoken.

It is important to note that in this case the expert team recommended the use of weak correlation. Weak correlation means that it cannot be ruled out that an otherwise disadvantaged person may not have a higher language exam and thus be unable to complete non-native language training.

The weak correlation makes sense when experts edit AI training datasets. In the D-ILA data model, we decided that in the case of individuals classified as disadvantaged target groups, there should be an individual who has more than one spoken language in addition to their mother tongue, but the proportion of such individuals should be very low in the 300-line AI training database. The question of negative discrimination has arisen: on what basis do we believe that disadvantaged people do not speak foreign languages? With this compromise solution, we did not exclude the occurrence of such cases, but at the same time we said that in our model only a low percentage of individuals with such characteristics can be included. Since the D-ILA data model is explicitly not representative, we do not make any claim that this is true at the societal level with the low proportion of such individuals. This approach applies only to the D-ILA data model.

Obviously, potential D-ILA users can make a different decision when constructing the AI training database. They can create an AI training database where up to 100% of individuals from disadvantaged groups will be able to participate in training that is not in their native language.

After agreeing to use the language in the data table of both the trainees and the training, we faced another difficulty in editing the 300-line AI training database – precisely because of the strong limitation of the spoken language. If all four participating countries enter their mother tongue (4 countries * 75 rows = 300 AI training series) and the language of the courses to match their mother tongue in their data sets, we essentially get 4 independent data sets, differentiated on a national basis. This approach would have run counter to the original objective of creating a meaningful data model at EU level.

The anomaly was resolved by approaching it from the training perspective. There was expert agreement that training, including soft skills training modelled within the framework of the project, is becoming increasingly international. There are two reasons for this. Firstly, it is supported by the EU, as it is a step towards a single internal market if training is made available at EU level. Secondly, machine translation programmes have reached such a level that translating training into any language can now be done very cost-effectively.

Both aspects pointed in the direction that although the language of the training can be specified, it is becoming increasingly irrelevant at EU level. (In line with our application objectives, we did not consider the interests of individual countries, which, for understandable reasons, may continue to focus on the importance of courses available in the national language.) Thus, during the development of the D-ILA data model, the partnership applied the solution that the language is included in the data table describing the characteristics of the training; however, this field was left empty in a significant part of the training. In practice, this solution meant that even if a mother tongue or other spoken language was provided for a given individual, who may be disadvantaged, during matching it could be assigned to training where the language of the training was not specified or matched the mother tongue provided for the individual.

Potential users of the D-ILA data model may make the same decision when designing their own data model, but may also make different decisions according to their own needs. For example, if the target group of your own data model is exclusively a domestic individual and the focus is also on the development of adult education within the country, then for each individual and for each training, the language data field will contain the mother tongue. If the target group is immigrants, language specification can be managed accordingly. The situation is again different if the model is designed for highly qualified managers of an SME, in which case it is advisable to create your own data model by specifying several spoken languages.

In the case of the trainees, a decision was made by experts to distinguish between mother tongue and spoken language, thus making the data model more precise. However, when testing the model, we found that the distinction between mother tongue and spoken language is irrelevant for the D-ILA data model, since spoken language and mother tongue were considered equal during matching. With this in mind, potential users of the D-ILA data model are advised to differentiate and use separate data fields to distinguish between spoken language and mother tongue only in justified cases. Only if importance is attached to the mother tongue when assigning to training.

The presented example only shows considerations related to the foreign language and the background of the decisions made by experts. There are 104 data fields in the D-ILA data model - although they partially overlap. It is also clear from this that compiling the data fields used in the data model is a serious task for professional experts. At the same time, it is necessary to think similarly to the presented example in order to be able to establish the necessity of a data field during the creation of the model. However, based on what has been described, it is clear that the selection and inclusion of a data field in the model does not create too much of an obligation, since the role of the given data field in the model can even be reduced by the conditions regarding the use of the data field. It is therefore a general recommendation, within reasonable limits, to include as many relevant data fields as possible in the model - as our team of experts did.

IV.2 Compilation of value sets

Compilation of value sets may seem to be an easier task than the compilation of data fields. Once you have agreed on the range of data fields to be used in your data model, the lists of values should be drafted quite quickly. However, in some fields of the D-ILA data model this was not the case, and we provide further explanation below.

In a few fields, the possible values are just a simple yes/no choice (e.g. if the training course offers a teacher/mentor for consultation or self-study is needed, or whether a participant is willing to learn independently). However, most of the fields require a whole set of possible values and the value sets have to be profoundly discussed. Here is an overview of data fields of the D-ILA data model where lists of values were set up:

Data field value set - Individuals	Data field value set – Training
Gender	Country
Country	NUTS2
NUTS2	Target group
Nationality	Language of the training
Target group	Occupation
Educational attainment	Accreditation
Occupation	Quality assurance
Language of the training	Delivery mode
Delivery mode	Supplementary services
Mobility	Educational attainment
Motivation	Certificate
Soft skills	Soft skills
Skills level	Skills level
Framework category	Framework category

Value sets were subject to internal discussions as there are two contradicting goals:

- they need to describe all variants of the variable, but on the other hand
- they need to be few and simple, in order to keep the database concise (each additional value multiplies the number of options and increases the size of the database) and easy to use for even non-professionals (i.e. intuitive).

Besides that, the requirement for the database to be used internationally brings additional challenges as practices on how some value sets are defined might differ among countries. The educational attainment or preferred target groups are good examples (see below).

Gender

There can be many gender identifications, but, based on empirics, most of them will be only marginally represented in a random population sample. Thus, the simple four choices seem to be sufficient for the practicality of the model: male/female/non-binary/no answer. Although the empirical research showed that gender is irrelevant for the training recommendation, in order to specify the target group, this data field was included.

Place of residence

It seems simple as there is an international categorisation of regions and everybody can select their place of residence. But the relation to training assignment generates problems:

- how far the person is willing to commute
- simple allocation of training taking place within the same region as is the participant´s residence does not solve the problem – the situation is diverse and individual: often, the time it takes to commute to another region can be shorter than to a distant place in the same region; in some cases commuting to another region is frequent and expected – e.g. in the Central Bohemia region, which surrounds Prague, the commute there is what almost everybody does and the provision of public transportation is very good, but it is not the case in other neighbouring regions. For a perfect solution, time and resources should be invested into an automated system that takes into account distances and public transport possibilities between the two places.

Education attainment

It is very useful information but it can be tricky in international databases. The national levels of education and related certificates are not easily linked to those of other countries. There are international classifications (ISCED, EQF), but these are not understandable to regular applicants for courses.

- for a database to be used only within one country, the relevant list of national terms for the education levels can be freely used,
- but in case the database is to be used internationally, the basic levels of educational attainment should be selected and described simply with understandable terms accompanied with some international classification (EQF and/or ISCED coding) to ensure comparability as well as clarity for common users.

In our data model we decided to use ISCED, ISCED F, and EQF for reference when creating value sets.

Funded by
the European Union

Level of skills

Usually, the level of skills is to be estimated by self-evaluation. Although, in case there is an internationally recognized online tool available (e.g. the language tests related to CEFR), it could be useful to incorporate the evaluation into the system, but it would require additional resources.

In the case of soft skills training, competence levels may have different value sets according to the framework, which must also be tracked at the ILA data model level. For example, in the case of foreign language competences, the value set has 6 elements (from A1 to C2), while in the case of DigComp there are 8 levels of each competence element. For other competences, 3 levels may be sufficient. This needed to be implemented in the data model.

Target group

Selection of target groups strongly depends on the given system (or country). Some target groups might not even exist or not be relevant in different countries or contexts. If, for example, the ILA scheme in one country is considered to improve the employability of all individuals participating in the labour market, employees represent a major target group here. In another country, it can be the opposite situation, where the focus is on public employment services that target unemployed or other vulnerable target groups.

Basically, there are three main principles to be respected when compiling data fields and value sets for a data model:

1. The discussion about what fields, on the one hand, and what related value sets, on the other, are to be implemented in the data model must run in parallel. The reason is that it is actually the concrete values behind each field that facilitate the decision making by making it comprehensible to the whole team.
2. Individuals and the training to be assigned to them must be defined with identical sets of values in order to make the subsequent matching possible.
3. While increasing the number of data fields and the range of value sets could improve the accuracy of matching, in a project it may be necessary to set a limit on the range and granularity of data for the sake of practicality.

IV.3 Generation of individuals' data

Compiling individuals' data to be applied in a data model may be done in various ways. In real life, the user would most probably make use of the existing (national) databases they have at their disposal and that fit their modelling goals. However, to a transnational project of a relatively small scale such as D-ILA-V4, no such data was available and the expert team had to generate data series on their own.

Partners from four D-ILA-V4 countries created and contributed equally, to a total of 900 individuals' data sets, by filling in a shared spreadsheet. Everyone was encouraged to complete it based on their own ideas and sources. In the end, the entire table was reviewed collectively and modifications were made in case an inconsistency was detected.

The project team discussed the extent to which the generated data should copy reality or be rather an attempt for equity. In the end, no strictly unified approach was established and each country

Funded by the European Union

approached the task in their own way. However, one principle was common - the data should approximate the diverse characteristics of real individuals.

Some partners, such as Czechia and Slovakia, derived a basic structure of their country's adult population first, in order to use it as a reference for creating a realistic list of hypothetical individuals. The Czech experts derived "quotas'' describing the Czech population from publicly available statistical data. This mainly concerned categories such as gender, age, a reference to the individual's region of residence (at level NUTS2), education attainment, and occupation. In the generated data, they tried to roughly follow the percentages. However, it must be noted that the final database did not strictly correspond to the population sample. For example, it is probable that among individuals searching for publicly supported training under the ILA scheme, the unemployed will be overrepresented. Therefore, in the database, the team increased the share of unemployed individuals more than it would be in a real representative population sample.

The Hungarian, Polish, and Czech partners generated the data series by rows. To generate "realistic individuals" it was very useful to think of an individual of a certain education level in a concrete life situation from which most characteristics could be easily derived. This included, for example, their existing skills and learning goals. In Slovakia, the experts generated most of the participant data by columns (i.e. data fields). The input was random; however, the result was controlled by respecting the most up-to-date statistical reports from the National Information System on Further Education. Thus, the shares of specific gender, age, education level categories, etc. matched the reality. A subsequent automatic check was applied to avoid inconsistencies between age, gender and target group value sets. All data fields were filled in except for the current occupation category and maximum volume of training to be undertaken. The team filled in these fields manually, and cross-checked the individual profiles row by row, looking for possible mismatches between the personal characteristics and the other data fields (especially those not covered by the national statistics and completely arbitrary, such as training goals, delivery mode preferences, etc.).

Altogether, 900 data series on individuals were generated by project partners manually, out of which 300 data series were used for training the AI. In the course of the subsequent testing process, it turned out that more data than 600 individuals would be necessary. Thus, new data series were generated, however, this time not manually but by using an algorithm. The generation of the new data series was based on a predefined data distribution matching the one of the original databases created manually by experts. It was done by columns, i.e. data fields.

Here is a list of instructions and reflections produced by the experts during the manual generation by rows:

- It is advisable to fill 3 rows at a time, because then it is easier to compare the values.
- It is possible to leave some fields blank - it corresponds to real-life databases where there probably will be some missing values.
- It is important to ensure that the values entered in the rows fully comply with the agreed value sets, neither codes nor the spelling should be different.
- The 3 budget fields must be filled in even if the value is 0.
- For the first few lines, we can think of a specific individual whose details we know approximately: a Not in Education, Employment or Training (NEET) young person who has acquired language skills in the school system but has not been able to find a job, has not

Funded by the European Union

undertaken training, and is under-motivated. They get a relatively large sum of public money for language training, which is likely to be good and useful, so that the individual can become more motivated to learn and enter the labour market.

- Check each row for logical inconsistencies, e.g. the NEET target group cannot be of a higher age, some target groups cannot be associated with employment.
- It is also necessary to avoid obvious contradictions. For example, it is not realistic for an individual with a doctoral degree to have no foreign language skills, or an individual living in the capital to be working in the agricultural sector (of course, this cannot be completely excluded).
- At the same time, it is advisable to be careful about biases, such as always placing individuals with higher education in the capital.
- In our case, there were 67 transversal skills providing the complete profile of an individual – but in the model there were only 4 options, so from the 67, the 4 most characteristic ones had to be selected – with the idea in mind that later the training will have to be assigned based on the 4 competences that have been entered here.
- It is advisable not to look at the ratios within a data field, but to concentrate on the rows at first. Once all the rows are completed, the ratios can be checked and then the values in the data fields can be changed.
- When it comes to individuals' goals, these could be random, since the data only represents a fraction of information about personal experience, interests, and contexts.
- Concerning the state's goals, a certain level of consistency between some characteristics, for example, age, place of residence, target group, and the goals, is welcome. This could reflect policy goals related to regions, target groups, and problems, such as e.g. regional priorities according to intelligent specialisations, age group (e.g. digital or health related skills for adults) or some national policy (e.g. lifelong learning - personal development and learning skills for all). Such an exemplary resource and possibly a simple tool for modelling could be presented as a very practical tool for most policies (since most, if not all, require skills development).
- In some countries there are policies introduced (e.g. in core curricula of higher education) establishing links between education levels and CEFR requirements. Therefore, at least theoretically, for some of the younger age groups we should expect more language competence at the given levels. This would be a very mild expectation - since learning language in schools may be ineffective.
- The variability of the amount of state funds available for a given individual matters little in terms of data, since it would be matched with the price of training (it's a simple numerical value comparison). The values can vary: funds may be accumulated and used, or additional entitlements can be made for specific groups or regions. At the same time, the values could be identical for all individuals to reflect a centralised governmental policy. However, in the end, the whole entitlement (government, employer, and individual) values would vary.
- In real life, individuals could have problems in profiling themselves. For example, DigComp may be difficult to grasp, while LifeComp categories seem intuitively understandable. Therefore, we can generate individuals' data to indicate that they have LifeComp skills and no previous training for that (as a result of informal learning) slightly more often. Some people may indicate digital skills or green skills with 0 hours on the same basis (e.g. Gen Z).

Funded by
the European Union

- It seems important to get the prices of training courses right. If we have values which are realistic, we can attempt to model the effects of various policy decisions in terms of the budget and allow governments to see what can be funded in a given period of time depending on the size of the allocation. Finally, we could provide an educated guess about “how much is too little”.

There were several issues raised during the generation of individuals’ data, which partners were not able to address in the framework of the project but which might be important for future users:

- If I want to indicate that an individual has a goal in developing a language competence, I can’t indicate which language it is. The same with the state’s goals - it may be relevant to develop a country’s official language in the case of immigrants, and a foreign language in the case of the majority.
- It would make sense to be able to indicate all language skills (listening, reading, writing, spoken interaction, and production) at once, while indicating only one of them would also be useful in many cases.

In conclusion, generating a database of fictional individuals proved to be quite a demanding exercise. At one point, joint reflections were made on the possible application of generative AI. There were attempts to use it for generating data series of individuals with pre-set parameters. However, the results were unsatisfactory as the algorithm was not able to respect all the requirements for each generated individual profile (N.B. The exercise was run in autumn 2023 using GPT Chat 3). That was the reason for not using AI for the generation of individuals’ data in the D-ILA model. (Nevertheless, it served well for generating course titles, for example).

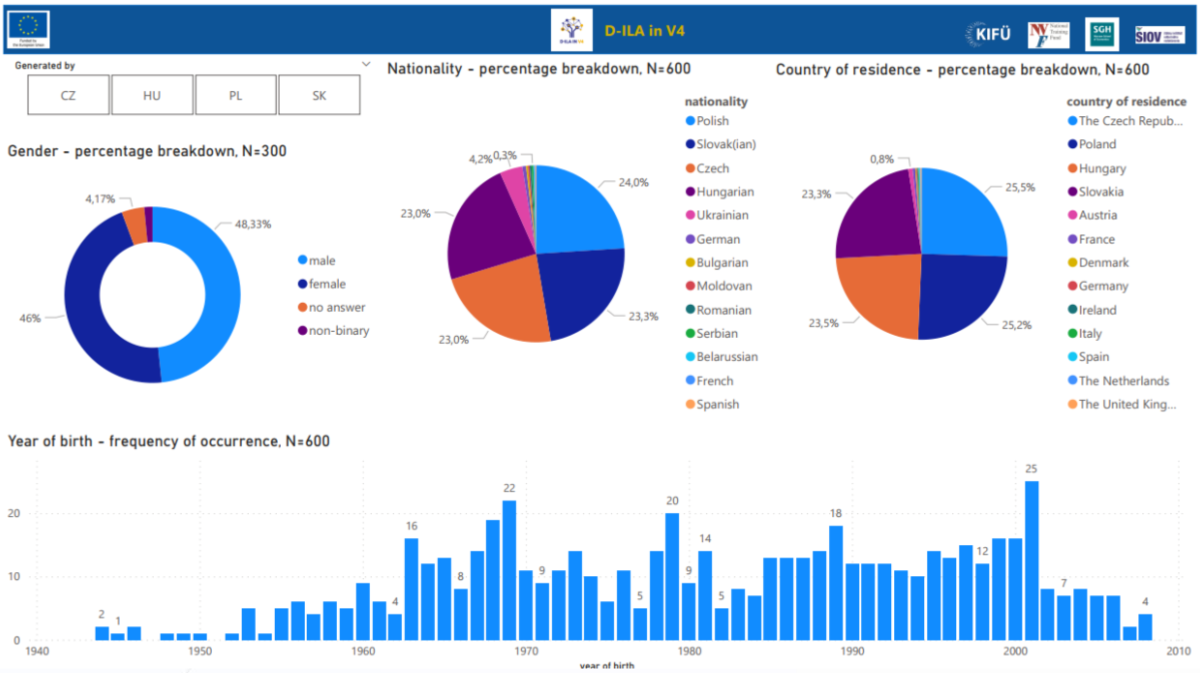
Here is an example of data series referring to one fictional individual:

P_ID	PL_P142
Gender	Female
Year_of_birth	1976
Nationality	Polish
Country_of_res	Poland
Place_of_res	Polska - Kujawsko-pomorskie
Highest_level_ISCED	Masters or equivalent level
Sector_of_highest_edu_att	Information and Communication Technologies
Highest_level_EQF	7
Target_group_1	Employed
Target_group_2	
Target_group_3	
Current_occu_cat	Information and Communications Technology Professionals
Native_language_1	Polish
Native_language_2	
Foreign_language_1	English
Foreign_language_1_level	C2
Foreign_language_2	French

Funded by
the European Union

Foreign_language_2_level	B2
Max_training_time_undertaken	14
Max_volume_of_training	100
Willingness_learn	Yes
Willingness_geo_mobility	Medium (within neighbouring region)
Willingness_occup_change	No
Primary_motivation	Personal interest
Budget_state_contrib	50
Budget_employers_contrib	100
Budget_private_contrib	0
1_skill_category	DigComp
1_num_of_hours	10
1_name_of_existing_comp	2.5 Netiquette (DigComp)
1_level_of_existing_comp	4
1_cert	No
2_skill_category	DigComp
2_num_of_hours	26
2_name_of_existing_comp	2.6 Managing digital identity (DigComp)
2_level_of_existing_comp	5
2_cert	No
3_skill_category	DigComp
3_num_of_hours	20
3_name_of_existing_comp	2.3 Engaging in citizenship through digital technologies (DigComp)
3_level_of_existing_comp	4
3_cert	Yes
4_skill_category	DigComp
4_num_of_hours	60
4_name_of_existing_comp	2.4 Collaborating through digital technologies (DigComp)
4_level_of_existing_comp	5
4_cert	Yes
Goals_individuals_1	3.1 Developing digital content (DigComp)
Goals_individuals_2	3.4 Programming (DigComp)
Goals_states_1	L1 Growth mindset (LifeComp)
Goals_states_2	4.2 Protecting personal data and privacy (DigComp)
Goals_employers_1	L1 Growth mindset (LifeComp)
Goals_employers_2	1.2 Evaluating data, information and digital content (DigComp)

As mentioned earlier, during the editing of the data describing a total of 900 imagined individuals, we were only able to ensure the correlations between the data per line. In other words, we had to find a solution for how we could create an overall picture of our 300 data series AI trainer and our 600 data series sample table. Our team of experts came to the conclusion that an interactive analysis application should be created with the help of Microsoft PowerBI, which enables the creation of an overall picture with the help of statistical statements. The figure below shows the statistical characteristics of the 600 data series sample[24]:



The interactive statistical analysis interface contains the following reports based on the data generated by the individuals imagined within the framework of the D-ILA project:

- Gender
- Nationality
- Country of residence
- Year of birth
- Level of foreign language
- Sector of highest educational attainment
- Current occupation category
- Highest level of educational attainment: ISCED
- Primary target group
- Secondary target group
- Tertiary target group
- Maximum training time to be undertaken [hour/week]
- Primary motivation
- Maximum volume of training to be undertaken
- Willingness for geographical mobility

[24] The data visualisation is available in its entirety via the official website of the D-ILA project: https://kifu.gov.hu/d-ila/ The downloaded .pbi data visualisation file can be opened with the free version of Microsoft PowerBI Desktop.

Funded by the European Union

- Available budget - state contribution [€] - frequency of occurrence
- Available budget - employers' contribution [€] - frequency of occurrence
- Available budget - private contribution [€] - frequency of occurrence
- Percentage of individuals having previously completed training by framework classification
- State's goal for individual's learning - percentage breakdown (2 goals per person)
- Employer's goal for individual's learning - percentage breakdown (2 goals per person)
- Individual's learning goals - percentage breakdown (2 goals per person)

Overall, it can be said that the visualisation of the D-ILA data sheets is very successful. Firstly, the expert team was able to continuously check the statistical characteristics of the edited data series on the fly. Secondly, during the validation of the data tables, the external experts were able to make sure of the data's realism.

We recommend those who want to use the D-ILA data model to create a visualisation of their data tables - whether they work with edited or real data.

IV.4 Generation of training data

This chapter provides some examples of interpreting textual description of training at the data level. First of all, the most important aspects:

- With the data, we can only approximate the diverse characteristics of real training courses. For example, two training courses that match in all the data can be completely different if they are taught by a very good teacher or a bad teacher. When filling it out, it should be expected that the conceived training will not be perfectly defined with the data provided. For example, increasing the number of data fields could be a solution, but we did not undertake this in this project.
- For the future, it is important to have a diverse range of training offers, which is what we should strive for when providing data series.
- 4 countries produced a total of 100 training data sets by filling in the shared spreadsheet. Everyone was supposed to feel free to fill it out according to their idea – and these examples – because in the end we will look at the whole table as one. If there is a training course that is significantly different from the others, there will be room for modification.

1. Training course description

This training focuses on a narrow area within green competences in a non-target group-specific way. It is of very high quality, which is confirmed by international accreditation and a high level of quality assurance. Since it covers a narrow area, 3 hours is enough to complete it. The training is carried out independently. The training has received accreditation for 3 years, that is, the content of the training has remained current for a long time. 50% of the course is practice, which indicates that there are playful exercises for 1.5 hours, e.g. based on waste photos, you have to decide which waste category the given object belongs to. Only minimal prior knowledge is required to complete the training. Since the training is highly validated, a digital micro-credential can be issued after successful completion. An increase in competence level is achieved in two green competence areas, which change from basic to advanced level as a result of the training.

Funded by
the European Union

1. Translating the description of the training course into the language of data

Generated by	EN
Name of training course	Environmental labelling on products
Unique training identification code	
Primary target group	No target group can be selected
Secondary target group	
Tertiary target group	
Language of the training	
Location of the training	
Related occupation	Unclassifiable occupation
Accreditation	Accredited by an international organisation
Quality Assurance	Internal QA systems
Number of notional learning hours	3
Delivery mode	Online at any time
Self-study needed	Yes
Teacher/mentor available for consultation	Yes
Supplementary services	
Possible only through independent learning	Yes
Price of the training course [€]	15
Last update of the training offer [year, month]	05.2022
Date of accreditation [year, month]	11.2021
Accreditation valid until	11.2024
The share of practical training within the total training period	50
Lowest educational attainment required to start the training	Primary education
Lowest EQF level to start the training	
Upon successful completion you will receive	Micro-credential

Upon successful completion of your EQF level	
Training evaluation based on feedback from participants	5
1. Framework category	GreenComp
1. ID and name of existing competence	Valuing sustainability (GreenComp)
1. Minimum competence level required for the training	Basic
1. Completion of the training provides an average level of competence	Advanced
2. Framework category	GreenComp
2. ID and name of existing competence	Promoting nature (GreenComp)
2. Minimum competence level required for the training	Basic
2. Completion of the training provides an average level of competence	Advanced
3. Framework category	
3. ID and name of existing competence	
3. Minimum competence level required for the training	
3. Completion of the training provides an average level of competence	
4. Framework category	
4. ID and name of existing competence	
4. Minimum competence level required for the training	
4. Completion of the training provides an average level of competence	
5. Framework category	
5. ID and name of existing competence	
5. Minimum competence level required for the training	
5. Completion of the training provides an average level of competence	

Funded by
the European Union

2. Training description

Training for the further development of green competence for engineering graduates, the data is set accordingly. For example, we are talking about hybrid training in English with a high number of hours. The price of the training course is therefore also high. The training also develops digital competences to achieve this goal.

2. Translating the description of the training course into the language of data

Generated by	EN
Name of training course	Developing the competence of environmental engineers
Unique training identification code	
Primary target group	Workers in jobs at risk from digitalisation/automatization
Secondary target group	Civil servants
Tertiary target group	
Language of the training	English
Location of the training	Hungary - Budapest
Related occupation	Science and Engineering Professionals
Accreditation	Accredited by an international organisation
Quality Assurance	Internal QA systems
Number of notional learning hours	60
Delivery mode	Hybrid
Self-study needed	Yes
Teacher/mentor available for consultation	Yes
Supplementary services	Group discussion and activities
Possible only through independent learning	No
Price of training course [€]	1680
Last update of training offer [year, month]	
Date of accreditation [year, month]	

Accreditation valid until	12.2025
The share of practical training within the total training period	20
Lowest educational attainment required to start the training	Master's or equivalent level
Lowest EQF level to start the training	7
Upon successful completion you will receive	Micro-credential
Upon successful completion of your EQF level	8
Training evaluation based on feedback from participants	5
1. Framework category	GreenComp
1. ID and name of existing competence	Valuing sustainability (GreenComp)
1. Minimum competence level required for the training	Advanced
1. Completion of the training provides an average level of competence	Expert
2. Framework category	GreenComp
2. ID and name of existing competence	Supporting fairness (GreenComp)
2. Minimum competence level required for the training	Advanced
2. Completion of the training provides an average level of competence	Expert
3. Framework category	GreenComp
3. ID and name of existing competence	Systems thinking (GreenComp)
3. Minimum competence level required for the training	Advanced
3. Completion of the training provides an average level of competence	Expert
4. Framework category	DigComp
4. ID and name of existing competence	4.4 Protecting the environment (DigComp)
4. Minimum competence level required for the training	3
4. Completion of the training provides an average level of competence	5

Funded by
the European Union

5. Framework category	DigComp
5. ID and name of existing competence	2.1 Interacting through digital technologies (DigComp)
5. Minimum competence level required for the training	3
5. Completion of the training provides an average level of competence	5

3. Training description

For a more popular training course, which covers 5 types of green competence areas for citizens – so it is not target group specific, the data given in the table has been selected accordingly. For example, the proportion of the theoretical part is extremely low. It is important that the level of individual green competences does not change, so participants are at the basic level both at the beginning and end of the training.

3. Translating the description of the training course into the language of data

Generated by	EN
Name of training	Developing a wide range of green competences for interested citizens
Unique training identification code	
Primary target group	No target group can be selected
Secondary target group	
Tertiary target group	
Language of the training	
Location of the training	
Related occupation	Unclassifiable occupation
Accreditation	Not accredited
Quality Assurance	Feedback from participants
Number of notional learning hours	8
Delivery mode	Online/Live
Self-study needed	No
Teacher/mentor available for consultation	Yes

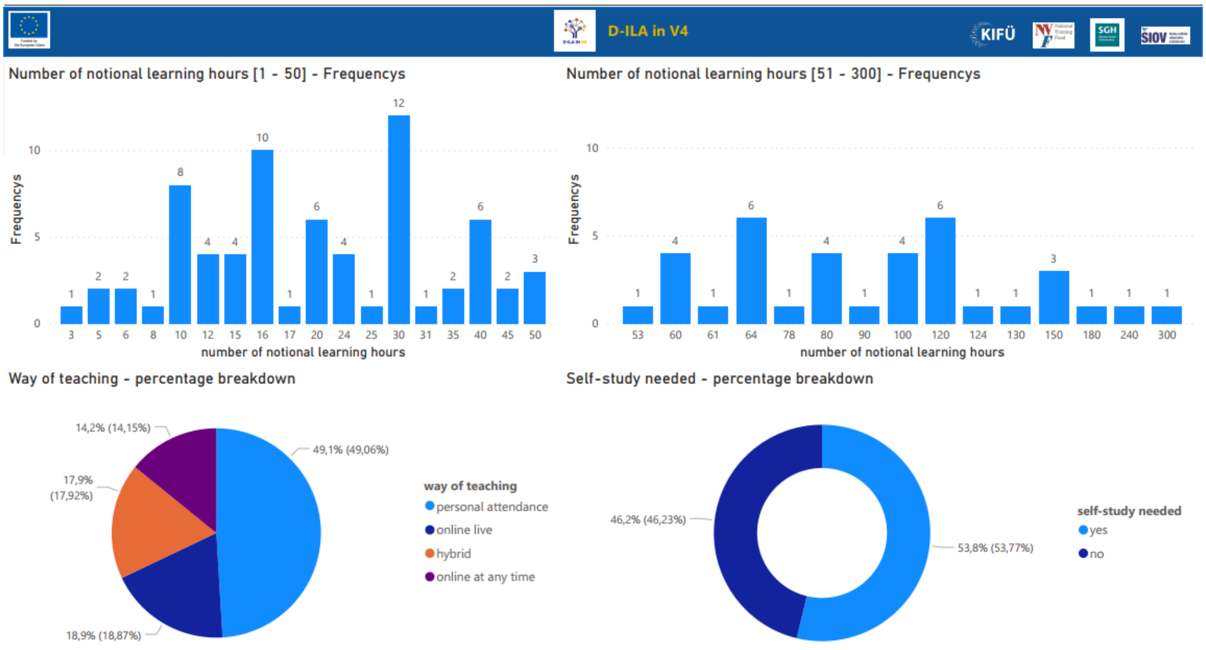
Funded by
the European Union

Supplementary services	Group discussion and activities
Possible only through independent learning	Yes
Price of training course [€]	9
Last update of training offer [year, month]	
Date of accreditation [year, month]	
Accreditation valid until	12.2024
The share of practical training within the total training period	85
Lowest educational attainment required to start the training	Primary education
Lowest EQF level to start the training	1
Upon successful completion you will receive	Not relevant
Upon successful completion of your EQF level	1
Training evaluation based on feedback from participants	5
1. Framework category	GreenComp
1. ID and name of existing competence	Valuing sustainability (GreenComp)
1. Minimum competence level required for the training	Basic
1. Completion of the training provides an average level of competence	Basic
2. Framework category	GreenComp
2. ID and name of existing competence	Adaptability (GreenComp)
2. Minimum competence level required for the training	Basic
2. Completion of the training provides an average level of competence	Basic
3. Framework category	GreenComp
3. ID and name of existing competence	Promoting nature (GreenComp)
3. Minimum competence level required for the training	Basic
3. Completion of the training provides an average level of competence	Basic
4. Framework category	GreenComp

Funded by
the European Union

4. ID and name of existing competence	Systems thinking (GreenComp)
4. Minimum competence level required for the training	Basic
4. Completion of the training provides an average level of competence	Basic
5. Framework category	GreenComp
5. ID and name of existing competence	Critical thinking (GreenComp)
5. Minimum competence level required for the training	Basic
5. Completion of the training provides an average level of competence	Basic

On the basis of the considerations described in the previous chapter, the editing was also carried out per data line for the training, because this was the only way to take into account the correlations within the data lines. It follows that in this case too, the visualisation[25] of the statistical characteristics of the data table containing the training was necessary. The figure below illustrates one of these solutions.



The following statements can be requested in the Microsoft PoweBI data visualisation file:

- Primary target group
- Secondary target group
- Tertiary target group

[25] The data visualisation is available in its entirety via the official website of the D-ILA project: https://kifu.gov.hu/d-ila/ The downloaded .pbi data visualisation file can be opened with the free version of Microsoft PowerBI Desktop.

- Country of training
- Language of training
- Accreditation
- Quality assurance
- Number of notional learning hours [1 - 50]
- Number of notional learning hours [51 - 300]
- Delivery mode
- Self-study needed
- Supplementary services
- Teacher/mentor available for consultation
- Upon successful completion you will receive
- The share of practical training within the total training period
- Number of courses in category
- Distribution of soft skills category
- Distribution of soft skills

The same can be said for the training courses: not only can the expert group continuously check the statistical characteristics of the edited data series on the fly, but during the validation of the data tables, the external experts could make sure of the realism of the data. We recommend it to those who want to use the D-ILA data model to create the visualisation of their data tables - whether they work with edited or real data.

IV.5 The matching exercise

As a reminder: the data table that we use to train the AI is created with the help of pairing. The pairing was carried out by our team of experts based on human intelligence and professional experience. The result of the pairing was that we assigned training to each person.

The experience gained during matching is summarised below:

- When it comes to matching, there is no ideal or single solution. In the D-ILA model, the objectives of the individual performing the matching to train the AI decides what aspects to consider and the order of importance. It is important to note that the order of the screening (filtering) criteria can influence the outcome of the matching.
- Competence is the strongest matchmaking aspect.
- If an individual has already completed a training course in a given framework, the level of the training completed (e.g. advanced) also applies if we want to develop another competence within a given framework.
- Several types of training can be selected for a particular individual. A possible approach is for the individual to attend all selected training courses. In this case, it is not advisable to provide two different training courses for the same competence development.
- In another approach, there is a logical “or” connection between the selected training courses. In this case, it is possible to provide 2 or 3 training courses for the same competence area, because then the individual will only be enrolled in one of them.
- In the D-ILA model, it does not matter much whether the individual, the state or the employer has indicated the competence to be developed. If, on the other hand, several individuals have

designated competences to be developed, it is advisable to prioritise. For example, if the employer provides the most support, the first step is to look for training for the competence it has identified. Only if there is no such training available, can the next step be to select training for the competence identified by the state.

- In some cases, belonging to different target groups are not mutually exclusive, in others they are. For example, a self-employed individual and an employee in the D-ILA data model belong to different target groups but may receive the same training. In contrast, an individual in a target group older than 50 years is much less likely to receive the same training than an individual in a NEET target group.
- The difference between the cost of the training and the ILA budget of the individual is less important for the matching. In the present model, it is assumed that if a very good match is found, it may be possible to attract additional funding - but this can of course be adjusted as required.
- Out of two training courses, we chose the one that develops several areas of competence for a given individual. This is not a top priority, only if both courses otherwise meet the higher level requirements.
- In many cases, the name of the training helps in the selection of training. This is because you can specify a free-text field that does not have a data field in the model.
- Upon the suggestion of AI experts, we removed from the D-ILA model the ability to prioritise selected training for a specific individual. During the matching, we see that this option would have been necessary after all.
- We thought that a given training could develop several competences at the same time, and we stand by this statement still. At the same time, one of the important experiences of matching is that competence No. 1 is given a higher priority in the practical implementation of matching, so we took this into account first. If we found suitable training based on this, the competence development effect of training No. 2 and No. 3 were considered as a lower priority.
- If we set a specific target group for training, the range of optional training will be very narrow. In our experience, it is not advisable to use the target group as a primary filter.
- Prejudices cause the same challenge with matching as with the production of edited data (e.g. fewer own resources available for the low-skilled). It is very difficult to avoid the problem; the inclusion of diversity can be the solution. If we recommend a variety of training courses for similar individuals, we do not condition AI that much for one approach. For example, if we ordered 100% cheap training courses for low-skilled people during expert matching, the trained AI would do the same in the future.
- A good example in the D-ILA data model is "people" who have fled Ukraine and are forced into low-status employment while having C1 level language skills. This is a good approach to diversity because AI is not conditioned to assume/learn that all refugees are poorly educated and poor.
- Multiple competence development needs for a given individual and multiple competence development services for a given training significantly complicate the matching. If a training course develops two competences out of 67 competences in the D-ILA data model, but only one of them has an explicit need, it is riskier to assign the training to that individual. The matching would be much easier if each training course focused solely on the development of one competence. This, however, contradicts our experience and expectations that transversal competences should be developed in a complex way.

- If we cannot find a training course that satisfies the competence development needs of the given individual, we can recommend similar training by filtering to the framework in the list of training. Thus, there is still a chance to choose training similar to the original need. Obviously, in the case of a competence that already belongs to a framework, the recommendation is very risky. For example, if an individual chooses a GreenComp competence for which there is no training, another GreenComp training course may be offered, but a language training course will hardly be offered instead.
- During matching, preference is given to training that satisfies an individual's multiple competence development needs at the same time.
- When matching, it is not advisable to think in statistical terms. You should not select a particular training course just because that training has not yet been selected. The best possible training should be selected for each individual.
- When matching, geographical location is almost impossible to consider. When matching in the D-ILA model, it is very unlikely to have two suitable training sessions and thus matching based on the location of the training.
- If several courses are equally suitable, the cheaper one should usually be chosen.
- The more often competences are identified as goals for improvement, the easier it is to find training for at least one of the identified ones. Where only one competence to be developed appears for the individual, and the training offer does not include training to develop that competence, it is difficult to offer an alternative training course.
- If the same competence development need is formulated by the individual, state or employer, it is advisable to treat it as a priority, i.e. to look for corresponding training in the list of training.
- The more precisely, with the help of more data fields, the characteristics of the individuals and training courses are specified, the more difficult the matching is, due to mutually exclusive conditions. A typical example is too detailed a target group definition.
- Life situation-based matching seems to be a good solution to bring together the competences of different frameworks, so a life situation can be the organising principle. This is an important proposal for the development of training courses to develop transversal competences.
- The efficiency of matching can be increased if the pool of participants is more homogeneous. In the D-ILA model, a wide variety of participants were generated. Compared to this, the originally planned 100 types of training often hinder the ideal matching.

V. Testing the D-ILA data model using AI

Concepts related to the professional content of the project and the D-ILA data model created within the project were presented in detail in the previous chapters. In this chapter, we discuss how we tested the operation and application possibilities of the D-ILA data model with the help of AI.

V.1 Policy scenarios tested using the D-ILA data model

In order to improve the quality of adult learning and increase the number of participants in adult learning, new financing concepts need to be developed, among other things. One of these new EU adult learning financing concepts is ILAs, for which the Recommendation of 16 June 2022 described the framework but left the details of implementation to national competence. Accordingly, Member States can adjust the eligibility and funding parameters to activate groups with traditionally low participation in adult learning and reduce the risk of deadweight losses. The specific target group will depend on the specific objective: increasing overall participation, increasing participation in training that meets labour market needs, integration of immigrants, parents returning after childcare, involving the elderly etc. When setting the level of public contribution, it should be borne in mind that differentiating the level of contribution according to the demographic characteristics of participants may lead to administrative complexity and thus create barriers to participation for groups with low participation rates in adult learning. The co-financing of activities (employer and individual contributions) and the definition of the learning content supported may vary from one Member State to another. All this shows that when implementing ILAs, many aspects need to be considered together to detail the policy measure.

In the framework of our Erasmus+ international partnership project, we explored the concept of ILAs from a data perspective. The cooperation of experts from the 4 Visegrad countries created the D-ILA data model to predict the possible impact of different policy scenarios. It is important to note that throughout the project, we worked with realistic but not real data created by the expert team through editing. In the baseline version, the expert team determined to the best of their knowledge which individuals and types of training should be included in the model, and they also determined which type of training should be assigned to which individuals (training database). In simple terms, modelling seeks to answer "what if..." questions, which in our case meant changing the input data along different policy scenarios and then comparing the original (baseline) version with the output generated with the changed input data. The method of analysing the impact of different policy interventions and the differences between them allowed us to test the D-ILA data model. The model is therefore methodologically tested, but the D-ILA data model needs to be populated with real data in order to make real use of the resulting figures.

Due to the call for proposals (Erasmus+ KA2 Adult Education), in this data model we have described learners and training courses using competence frameworks describing transferable skills (CEFR, DigComp, FinComp, EntreComp, LifeComp), and we have also tried to imagine training courses that develop skills belonging to several different frameworks at the same time (e.g. digital and green competence).The amount of funding (€400.000) did not allow for AI development, we were only able to demonstrate the relevance of modelling using existing solutions. By presenting professional concepts developed and tested by an international team of experts, we aim to highlight the usefulness of the D-ILA data model.

Some of the professional concepts tested were aimed at providing additional resources for the participation of a selected target group in adult learning. The D-ILA data model allowed the target group to be defined in several ways:

- by the competences preferred by the participants
- by the highest level of education
- by participants' life situation category

The D-ILA data model was able to support targeting according to the interests of the learners. A good example of this is the significant increase in the public contribution for participants who wanted to participate in green skills training. As no other parameters were changed, the results obtained after running the AI showed the impact of the professional concept.

A similar approach was taken when our policy scenario was that the state preferred to develop digital competences rather than green competences. (It is important to note that in the D-ILA data model, a strong emphasis is placed on training that develops several competences - for example green and

Funded by
the European Union

digital competences - within one training course.) Our policy approach of allocating additional public resources to the development of digital competences also differs from the previous one in that we have incorporated a significant increase in employer contributions for digital competence development. This is based on the premise that while green skills development is less relevant for an employer, digital skills development is in the employers' core interest - i.e. employers, not just the state, want to actively participate in the financing of digital skills training.

In another policy scenario we tried, public resources were spent in a more focused way. Only and exclusively, those who aimed to develop digital competences received public funding. At the same time, the policy scenario did not exclude non-digital skills training from the system, as it did not eliminate financial contributions from employers and individuals.

Target groups can also be formed by differentiation according to the highest level of education. Compared to the baseline, the D-ILA data model has been used to test the case where those with tertiary education are not subsidised by the state - assuming that they have the resources to participate in adult learning without state support. In this case, our professional approach did not imply resource extraction, as the public resources taken away from those with tertiary education were fully distributed among those without tertiary education.

The third way of creating a target group – provided to users by the D-ILA data model – was based on the classification of participants into life situation categories. In this professional concept, the training of NEET young people is financed by other means outside the ILA framework. The public support initially signed for them was therefore distributed among the unemployed participants and the effectiveness of the support was tested this way.

The D-ILA data model was also tested in a case where significant additional public resources were allocated to the most deprived compared to the baseline, i.e. here the focus was not on reallocating public resources. Here too, we assessed the need according to the living situation category of the participants.

In addition to policy scenarios based on target group differentiation, we also tested economic cost-benefit policy scenarios using the D-ILA data model. One of these was to make it a condition of state aid that the learner should also contribute to the financing of their own training – and this was compulsory. In other words, the condition for state aid was an individual contribution. After various considerations and lengthy discussions, it was decided that the compulsory contribution should be 20%. In another case, we tested the effects of a compulsory contribution not only from the individual but also from the employer. In addition, there was a policy scenario where we waived the compulsory co-payment for participants in certain target groups (in previous versions we gave additional state support on a means-tested basis, in which case the cost reduction was for the most deprived).

The policy scenarios presented are based on different guidelines. However, the applicability of the D-ILA data model is further enhanced by the fact that a policy scenario can be tested with different parameters. By not only modelling the 20% co-payment in the policy scenario described above, but also calculating the result with different % ratios, we can compare the different impacts and, if necessary, make a better decision using the D-ILA data model.

The EU concept of ILAs assumes that funding is also an incentive. This was also tested using the D-ILA data model: only participants who did not have personal learning objectives received public funding. The other "motivated" participants could only use individual or employer resources.

Funded by
the European Union

The policy scenario of giving accredited training a greater role in the D-ILA data model is intended to improve the quality of optional training. This has been achieved by reducing the price of accredited training, partly departing from the ILA concept, thereby increasing the likelihood that cheaper training courses will be offered to a greater number of participants in the D-ILA data model.

This example illustrates that even an unconventional approach can produce good results – it is just that this should be verified by modelling prior to implementation. (The D-ILA data model is of course not a complete modelling system in its current form and further development is needed to apply it in real-life contexts.)

Another interesting approach is the policy scenario that favours the provision of a particularly broad range of training courses. In this case, we have examined the effects of expanding the supply of training courses by offering training that is relatively inexpensive and can be relevant to a wide range of individuals.

The next policy scenario intervened in the supply of training courses by modularising them and analysing its effects. This aimed to halve the time spent on learning, reduce costs, and reduce the number of competences that the course develops. The feasibility study carried out earlier in the project confirmed the validity of this concept, as there is a growing demand for short and focused modular adult learning courses to replace traditional adult learning courses.

One group of policy scenarios created by our team of experts focused on the fact that state aid was not linked to a target group or any other condition, but was essentially treated as a benefit as a citizen's right. Thus, the policy scenario set a baseline where all participants, without exception, received the same amount of public assistance, while the amount of employer subsidies and individual contributions remained unchanged.

In the next step of the modelling, we made changes to the input data such that the ILA concept was reduced to exclusive state support, i.e. all employer and participant financial contributions were set to 0 in the D-ILA data model. With this occupational concept, we modelled the effects of a significant resource withdrawal. In particular, we paid attention to how this scenario impacted the NEETs and the low-skilled. We also tried a less severe version of resource extraction, where we "only" reduced the contribution of participants to 0, while imposing a mandatory contribution from employers.

The issue of the quality of training has been raised several times in the policy scenarios. In addition to state funding as a citizen's right, we have also tried out the option whereby the same amount of state funding for all can only be spent on accredited training. This is a significant conceptual change since, in the basic version, state funding was still freely spendable, whereas in this case the funding provider's constraints are also reflected in the D-ILA data model.

During the testing of the D-ILA data model, we also encountered some limitations that did not allow for the analysis of a certain type of policy scenario. Typical cases are those where the logical sequence of input data setup – AI run – output data generation and analysis is changed. This occurs, for example, when we want to determine the level of financial support in the policy scenario depending on the training course chosen.

In summary, the D-ILA data model has been shown to predict the impacts of a wide range of policy approaches, helping to inform policy decisions. We are convinced that this approach could be of interest in many areas of education and training – and if so, targeted AI development and improvements in data quality could lead to an even more accurate model.

Funded by
the European Union

V.2 Indicators developed to evaluate the results of AI runs

To be able to evaluate the various policy intervention scenarios, several indicators have been selected. This allows comparing different types of interventions using a set of common metrics. However, the assessment of each individual scenario goes beyond these indicators or requires a more nuanced or detailed approach.

A perfect model assigns courses to individuals reflecting real-life choices - in such a case the model prediction would reflect adult learning participation rates (within the ILA scheme and related training database), its prediction would be exact. Our model takes into account variables such as budget, availability of time, learning goals or preferred mode of learning (as described in the previous sections).

The following set of indicators has been developed for assessment of the scenarios results:

A. Number and percentage of individuals for whom AI has not provided a training offer

Being able to recommend or provide a learning opportunity is a metric for success for career counsellors as well as for the D-ILA model. The metric does not provide information on the quality of the matching or number of training courses proposed for a given individual.

Many of the policies for adult learning strive to increase the number and percentage of the population participating in adult education and/or lifelong learning. Using this metric we can evaluate how many more (or less) training opportunities can be recommended to an individual before and after the intervention. As the model includes data related to relevant barriers for adult learning participation (e.g. budget, time constraints), these indicators can be used to evaluate if a scenario aimed to alleviate these barriers will be effective.

Calculation method:

- The number of individuals for whom AI has not provided a training offer:

$$Sample\ size$$
$$-\ the\ number\ of\ individuals\ in\ the\ sample\ assigned\ at\ least\ one\ training\ offer$$

$$(\textstyle\sum_{x=i}^{n} \quad x_i = x_1 + x_2 + x_3 + \ldots + x_n) - (\textstyle\sum_{y=i}^{n} \quad y_i = y_1 + y_2 + y_3 + \ldots + y_n)$$

$$where:$$

$$x \ = \ individual\ in\ the\ sample$$

$$y \ = \ individual\ in\ the\ sample\ with\ at\ least\ one\ training\ offer$$

$$n \ = \ upper\ bound\ of\ summation$$

$$i \ = index\ of\ summation$$

- The percentage of individuals for whom AI has not provided a training offer:

$$\frac{The\ number\ of\ individuals\ for\ whom\ AI\ has\ not\ provided\ a\ training\ offer}{number\ of\ individuals\ in\ the\ sample}$$

B. Number of training courses offered (total and per individual)

Funded by the European Union

A higher number of training courses recommended allows individuals to choose from a wider range of training courses, thus giving them more freedom of choice. The indicators allow for comparing the results of two scenarios in terms of the success of the training recommendations.

The indicator does not provide information on the quality of the matching, so we do not know how suitable the recommended training is for each individual. The indicators are sensitive to the budget parameters, but also depend on the supply of training (number of courses, their length etc.). These indicators can be used to evaluate scenarios which affect both the supply and demand side of the training market.

Calculation method:

- The number of training courses offered:

$$The\ sum\ of\ training\ courses\ offered\ to\ all\ individuals\ in\ the\ sample$$

- The number of training courses offered per capita:

$$\frac{the\ number\ of\ training\ courses\ offered}{number\ of\ individuals\ in\ the\ sample}$$

C. Expected cost of training funded (total and per individual)

The entitlements available in the ILAs can be used to different extents. For the training to take place, a number of requirements need to be met, including individual motivation, but also limitations of time, money or distance. For the policy makers however, a significant issue is the cost of the ILA scheme, and the money that needs to be put aside or the moment at which it is spent (e.g. first year/second year of the scheme). This indicator can be used to assess costs of scenarios with different funding parameters (e.g. 300€ or 500€ per individual per year) based not on simple arithmetic, but also the prognosed spending rate of individuals.

The expected cost of training indicator is based on several assumptions: (1) the prognosed training offer reflects choices that individuals would make; (2) individuals would choose only one course (even if they could finance two training opportunities from the budget); (3) there are no limits related to the supply of courses (e.g. the number of trainees cannot be too small or too big).

Calculation method:

- The expected total cost of training:

$$The\ sum\ of\ costs\ of\ the\ first\ (best\ fit)\ training\ course\ offered\ to\ all\ individuals\ in\ the\ sample$$

- The expected cost of training per individual:

$$\frac{the\ expected\ total\ cost\ of\ training}{number\ of\ individuals\ in\ the\ sample}$$

D. Number of training offers broken down by competence groups

Funded by
the European Union

In the D-ILA data model we use competence frameworks for tagging training offers. A training course can be classified as being related to up to five transversal competences from one or more of the frameworks (for example: one course can combine skills related to DigComp and EntreComp). The metric shows the total number of training courses recommended broken down by frameworks and is presented as a table.

This indicator can be used to assess the effectiveness of various scenarios aimed at increasing the provision of training related to selected types/groups of competences (e.g. key competences, digital competences). In our model, only competence groups that are related to soft skills and European competence frameworks are being covered with this indicator.

When used on data based on the number of actual courses taken, this indicator may be useful to evaluate which types of competences are often jointly taught/learnt. Such knowledge may indicate complementarity of competences (for example, increased propensity to provide them jointly - in one course, and/or a practice in exercising them together) or can provide information about trends and dynamics in learning provision.

Calculation method:

For each training offer recommended, the framework combination is being identified. We then count how many training courses are characterised by a given framework combination.

The example below shows that most training offers included only GreenComp, a mix of EntreComp and FinComp or were language courses (CEFR).

Num. of training courses [pcs]	1_framework_category	2_framework_category	3_framework_category	4_framework_category	5_framework_category
233	GreenComp				
215	EntreComp	FinComp			
194	CEFR	CEFR	CEFR	CEFR	CEFR
159	LifeComp				
144	DigComp				
124	LifeComp	LifeComp	LifeComp	LifeComp	LifeComp
111	EntreComp	FinComp	FinComp		
103	DigComp	DigComp	DigComp	DigComp	
…	…	…	…	…	…

Funded by
the European Union

E. Number and percentage of individuals who have been offered training by competence groups

By analysing the types of competences related to training offered to each individual, we can learn which competences match well with individual needs and limits (which is modelled based on the data provided). The metric shows how often each competence group has been offered to individuals.

This indicator helps in evaluating if an intervention aimed at increasing the development of a specific type of skills (e.g. green or digital competences) is effective.

Calculation method:

For each competence group the sum of individuals who have received at least one training offer with this competence is taken. Two values are used:

- Number of individuals undergoing training under a specific framework [person]
- Share of individuals participating in training under a given framework in the sample [%]

The results are presented in a table, as in the example below:

Framework category	Base scenario	Analysed scenario	Base scenario	Analysed scenario
CEFR	116	116	19,3%	19,3%
DigComp	310	323	51,7%	53,8%
EntreComp	296	304	49,3%	50,7%
FinComp	212	233	35,3%	38,8%
GreenComp	160	156	26,7%	26,0%
LifeComp	295	277	49,2%	46,2%

F. All competence level changes available through the recommended training

Data fields describing training characteristics include "minimum competence level required for the training" and "completion of the training provides an average level of competence". Thus, from these two data fields, the added value can be calculated for each training course according to the competence group in question, i.e. the difference between the level expected at the start of the training course and the level achievable at the end. Since a training course can be described with 5 types of transversal competences in the D-ILA data model, the added value of a given training course can be interpreted according to a maximum of 5 transversal competences. Level progression is not meant per framework but per domain within a framework.

Important! It can be misleading to compare the sums of changes because they are closely linked to the number of courses offered (of which, often only one will be selected), in many courses there are double indications of increase in levels (e.g. an increase in competence level in speaking and writing in CEFR). To improve the indicator, we see a need to counteract the bias related to the quantitative approach and attempt to show a more nuanced construction. We propose to use the average, or the median, value from the courses offered or select the best fit course for the indicator. In all cases, further testing and development of this indicator is recommended in further implementation.

Method of calculation:

To calculate the change in competence level due to training, we determine the difference between the "minimum_competence_level" and "existing_competence" data fields of the D-ILA data model. If a given training course develops several competence groups (maximum 5), the difference between the two data fields is calculated several times for that course. To determine this indicator, changes in all competence levels across all offered training levels are added up per framework. The indicator is not interpreted for a specific individual during the analysis. Competence levels expressed as text (e.g. basic, advanced) are replaced with numbers so that the numerical score can be calculated. As a given training course may cover more than one framework, it will count towards more than one framework. Level progression is not meant per framework but per domain within a framework.

Important! In interpreting this indicator, it is not acceptable to compare the levels of different frameworks, because one level does not mean the same thing in two different frameworks (because a framework can have a different number of levels: 3, 6 or 8 levels).

Sum of changes in all competence levels within a given framework for all AI-recommended training [score].

Framework category	Base scenario	Analysed scenario	Change
CEFR	1036	1254	121,0%
DigComp	2083	2934	140,9%
EntreComp	2607	3154	121,0%
FinComp	569	750	131,8%
GreenComp	396	860	217,2%
LifeComp	1412	1415	100,2%

It may be that for some scenarios, not all the indicators will be relevant. Whereas for other scenarios, different, specific indicators would be needed. For example, policy scenarios aimed at supporting developing only selected types of competences may limit the evaluator interest in selected types of competences developed. While policy scenarios aimed at increasing the uptake of high quality training offers, may require introducing a new indicator reflecting the percentage of individuals choosing training that is accredited to certain standards (or meeting other qualitative criteria).

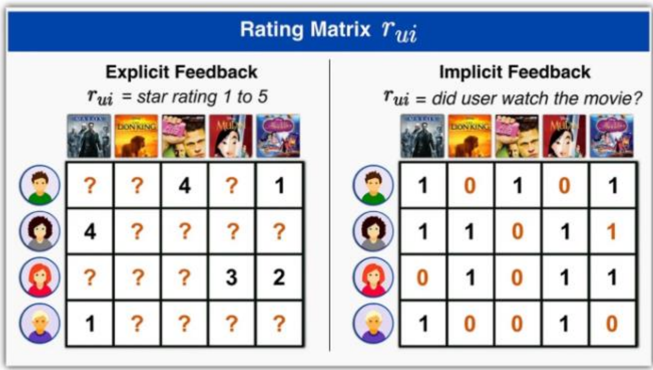
Funded by
the European Union

V.3 How AI works

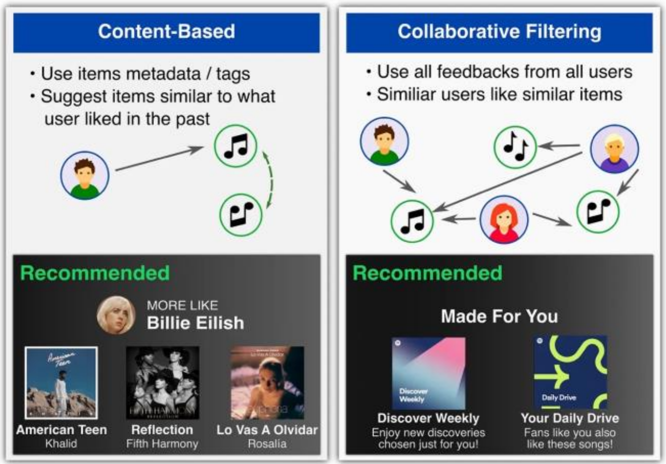
The project focused on the development of an AI-based course recommendation system in a hypothetical scenario. In order to ensure quality data, experts in the field produced both course details and participant profiles and then performed the matching between datasets. Using these selected matches, a personalized recommendation algorithm was created for new individuals.

The basic ideas of recommendation systems are as follows:[26]

- Algorithms that aim to help users in decision-making by providing personalised suggestions
- Based on the type of data (explicit vs. implicit feedback) there are different algorithms to use
- Based on the feedback we create a user-item rating matrix
- Question of 'cold-start': new item/user appears



- Content-based: if a user selected item x, we can recommend item y based on the items' similarity (Can be used with clustering algorithms)
- Collaborative filtering: use the history (feedback) of all users (cluster or factorise the rating matrix to predict unobserved item-user pairs)
- Hybrid approaches use both



The recommender created within the project implements a course recommendation system that utilises implicit feedback data to suggest courses to participants based on their interactions

[26] Source: https://towardsdatascience.com/recommender-systems-a-complete-guide-to-machine-learningmodels-96d3f94ea748

Expectations of the recommender:

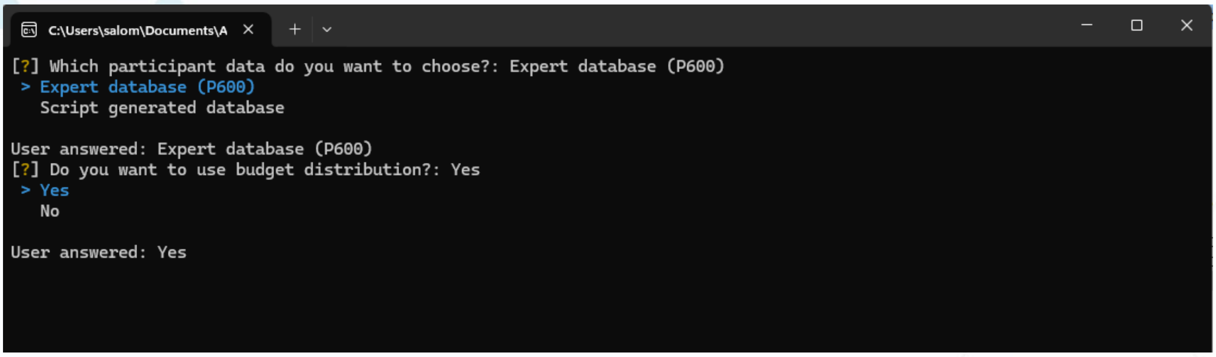
- processes participant data
- extracts preferred course categories
- suggests courses aligned with participant goals and available budget
- customisable as you can modify the recommender's inputs through command-line prompts

Steps:

1. Takes a participant ID as input and finds the most similar participant in the training data
2. Generates recommendations by calculating user-item interaction scores, and returns a list of recommended courses with their scores
3. Extracts the participant's ranked wishes (participant/employer/state)
4. Matches and filters courses that align with the participant's required skill categories
5. Filters recommended courses based on the budget constraints of a participant. It calculates the total budget available for a participant and selects courses that fit within that budget
6. Writes the resulting participant-courses matching in an excel file

With the AI-based solution developed within the framework of the project, it is possible to generate multiple outputs. After launch, two questions will pop up, they can be selected using the arrows on the keyboard, after which you can move forward by pressing Enter. Accordingly, the programme generates an excel file and saves it in the same folder as the exe file. The file will also include choices in its name, making it easier to navigate through the .xls results participants get. The structure of the dumped file: one column contains the participant IDs and the next columns contain the courses.

The figure below shows the interface for running the AI. Since no AI was developed in the project, we strove for minimalism - as can be clearly seen in the simplicity of the interface.



During the testing of the AI, the question arose whether there is a priority order between several training courses (there can be 27) assigned to an individual in the results table, and is training with a lower sequence number recommended? Or does the algorithm treat the proposed training courses equally, so the common feature of the training courses in the list is that they have passed the compliance filter?

The recommended training courses are sorted, so the one with the smallest "serial number" best suits the participant, but all those that are included in the list have passed the filter. It can happen that there are a lot of training courses because, if the participant has a large budget, the algorithm will recommend it until this amount is full, assuming the algorithm finds a suitable course at all. Most likely, at the end of the list, there will already be less relevant courses.

When running an AI solution, the software asks two questions, one of which is: *Do you want to use budget distribution?* During testing, both yes and no options were tried and the results obtained were compared. According to our preliminary expectation, where the training recommendation rule is

looser, fewer people will be left without training options. The following indicators were used to test the yes/no option:
600 individuals + yes Select option:

- 1,867 training courses offered
- 41 individuals did not receive a recommended training course

600 individuals + no options have been selected:

- 2,032 training courses offered
- 147 individuals did not receive a recommended training course

The results obtained can be interpreted according to the following principle:
The difference between the two options is that they sort courses differently, but one of the two sorts cannot be said to be "looser" than the other. If we choose yes, we prioritise the skills that the source which contributes the most (state or workplace) wants the individual to learn. If we choose “no”, we always prioritise the skills that the individual wants to learn.

Accordingly, there may be certain trends, e.g. the price of training resulting from each "choice". If, for example, it is generally true that individuals want a more expensive training course than the state or the workplace, then even one of the most recommended training options may not fit into the budget. On the other hand, there may be differences in who has what budget. For example, there is bias where the state/workplace does not contribute at all, because there are no expectations in the data generated by experts as to what they "think" the individual should learn. In such a case, if we run "no", a smaller amount of resources is available from the outset and may not be sufficient to cover the cost of the recommended training. And if we run “yes”, there is a bit of bias by default, since we prioritise the training course for which (all) resources have been provided. It is also possible that due to the more skills to be learned (sorted), the participant will have more recommended courses, so there is a better chance that there will be one that fits into the budget. If an individual only wants to learn digital competence, then obviously the system only recommends them and looks for suitable training options. If Green skills appear next to it, the system recommends several of them, so more can be included.

V.4 Results of AI runs

The results of the simulations performed using AI in the D-ILA project (in brief: the AI runs) are performed on synthetic data, therefore it would be best to see them as prognoses for a virtual state. For this state, the model prognosis can be seen as very accurate, however, because the state is not real - we can only rely on statistical tests to attest this (it cannot be validated with real data observed after the prognosis is done). As such, these results should not in any case be understood as real life recommendations for the effectiveness of any real policy scenario. To be useful for a real state or a company, the model needs to be adjusted to the population that is to be modelled.

The AI runs provide a proof of concept: the modelling is possible and gives reasonable results (i.e. results that passed a human ‘sanity check’). However, the actual changes observed as a result of the prognosed interventions should be treated with caution.

The model usability has been tested on cases of 20 proposed policy interventions related to adult learning support using ILAs. For each AI run, a separate description of the proposed concept of intervention, the changes implemented in the data, the outputs of prognosing, and evaluation of the results has been developed. The interventions have received short titles and are presented below in thematic groups:

Group 1. Scenarios supporting development of selected competence groups:

- Support of ICT skills
- Increase GreenComp budget
- Shift the level of digital competences of citizens

Group 2. Scenarios focused on selected target groups of learners:

- Excess funding for the unmotivated
- Those with a higher education attainment do not get state money (discrimination of the educated)
- Concentrating ILA support on the unemployed only (other groups are supported by other policy measures)
- The state doubles support for those in need (selected target groups)
- Comparison of PL and HU target group results

Group 3. Scenarios related to modelling the contribution schemes:

- Introduction of a compulsory individual contribution
- Individual and/or employer contribution (obligatory 20% of co-funding of state aid)
- Target group specific co-payment obligation
- Only state contribution
- Mandatory employer co-financing
- Individuals who choose non-accredited training courses receive less state aid

Group 4. Scenarios related to supply or quality of the training offer:

- Incentivising choice of high quality training
- Targeted expansion of the training list
- Splitting long courses into two
- Reducing the quality of training courses (negative scenario modelling)
- The training offer is limited to accredited courses only
- Doubling the number of courses and halving the price

The results have proven that the model is effective for prognosing scenarios focusing on changing the size of the entitlement, as well as the costs of training. It allows for prognosing targeted support of specific groups (e.g. by increasing their entitlements) and competence development (e.g. by introducing changes to training costs, reflecting subsidies). Other parameters may also be manipulated, however, some of them - such as level of educational attainment - have been treated as fixed. The indicators proposed allow for observing the externalities of selected scenarios (for example, an increase in the number of DigComp training courses offered may coexist with a decrease in the number of LifeComp training courses offered). The AI runs have shown that manipulating the training parameters is also possible. Although expanding the training list (adding new training courses) requires retraining the model, which is much more time consuming and was not possible in the project.

After that, let's look at some more interesting AI run result evaluations. (The measurement protocols - including the results of the AI runs - are available in the electronic appendix of this methodological guide on the project's official website.)

1. Introduction of a compulsory individual contribution

The aim of the simulation is to inform the decision on whether it is appropriate to introduce a mandatory contribution to education by the individual. The options considered are a state contribution for all without distinction and a state contribution for all but with a 20% contribution from the individual. Some of the political representatives in the country see these funds as an incentive for citizens willing to complete their education and invest their time and money in it, although this measure risks excluding the less well-off part of the population from adult education.

In particular, it would be interesting to know whether individuals would be willing to invest the necessary EUR 400 of their own money. However, this cannot be determined from the data model; it would be necessary to supplement the information from, for example, empirical research on this topic, or the experience from the ongoing Czech e-course shop run by the Ministry of Labour and Social Affairs, which has set up similar training conditions.

In our model, presumably due to the automatic allocation of relatively large amounts of training money (2000 EUR from the state and 400 EUR from the citizen's contribution) to each individual in the database, almost all individuals in the database participate.

In terms of the fields in which the courses have increased the level of competence of the citizens, the two variants are very similar. Most of the new skills acquired were in the area of business skills, followed by digital skills. Slightly lower skill increases are observed in Life Comp and language skills. Financial skills development was the least supported. The most significant difference between the scenarios is in the case of green competences, which were selected rather less overall, but significantly more in the second option with mandatory contribution from the citizen.

In this case, assessing the efficiency of the invested funds from the point of view of the state is difficult. The total number of courses offered was slightly higher in the option with a compulsory citizen's contribution, in the model situation this would mean more educated individuals with the same state contribution, and the option with a citizen's contribution would be slightly more profitable. In reality, however, some individuals would not participate in this option precisely because of the need for a contribution, and it is likely that it would be individuals with lower incomes and less education who would remain excluded from adult education.

2. Those with a higher education attainment do not get state money (discrimination of the educated)

Surprisingly, compared to the base version, the application of the professional concept significantly increased the number of individuals to whom the AI could not offer training courses. At the same time, the number of courses offered also decreased. Meanwhile, the total cost of the offered courses did not change essentially (indicator number 3).

Based on the 4th indicator, the order of the popular courses has also been rearranged. DigComp and courses that develop foreign language competences came to the fore, while EntreComp and FinComp courses were relegated to the background. This phenomenon might be explained by the fact that those with lower qualifications turn to more practical training if they get extra funds. Meanwhile the withdrawal of funds from those with a higher-education attainment results in a decrease in the number of recommendations for more theoretical training.

The analysis of changes in the level of competence (indicator number 5) also shows interesting results: in the case of all competence frameworks, the level of increase in the level of competence due to the

Funded by
the European Union

training decreased. Looking at the total population of 600 individuals, the transformation of the support system according to the professional concept reduced the increase in the competence level achieved by the training course at the system level. In other words: on the basis of the D-ILA data model, the redirection of state aid towards those with a lower educational attainment does not increase the level of competence at the population level. Lower-educated individuals are less able to increase their competence level despite the additional resources and higher-educated individuals are less able to increase their competence level due to the withdrawal of resources.

However, this image is extremely deceiving. The decrease in level rise can simply be caused by the fact that fewer individuals received fewer training offers, and this is the reason for the phenomenon shown in the data.

Indicator number 6 is in line with what was described earlier: in more or less all cases, the number of individuals participating in training for a certain area of skills decreases - with the exception of green skills. In this case, the reallocation of resources does not seem to have changed the number of participants. As much as the number of participants with a higher education attainment could decrease, it could increase among those with a lower education.

3. Incentivising choice of high quality training

The proposed evaluation of the AI run results requires a general disclaimer: the results are only as good as the data – the better the data meets the realities of a given country, the more accurate the results of the modelling.

Modelling the proposed intervention provides the following insights:

- The intervention increased the range of possible choices for individuals (by 0,37 courses per person from an average of 3,34 to 3,71 courses per person) and, of the training offered, the percentage of high-quality offers increased by 11,1 per person. (from 34,9% to 46%).
- Supporting high-quality training in the way proposed has yielded an over 11,5 per person increase in the percentage of recommended high-quality training offered (from 68,2% to 79,7% of individuals being offered at least one high-quality course).
- The recommendation of a given training course and the actual choice made by individuals may not be identical, but it seems safe to assume that the budget constraint is one of the most relevant barriers for choosing a specific training course.
- As quality assurance is expensive and the costs will ultimately be paid by individuals from ILAs, it seems that an intervention focused on supporting high-quality and on reducing the costs for individuals makes sense. The costs of such intervention can be calculated based on the number of individuals receiving a high-quality offer and their willingness to choose one of these training courses: in the scenario, 478 received at least one high-quality offer depending on the course uptake rate, the cost of the intervention varies between €21.000 and €57.000.
- The negative aspect of this intervention is the ratio of high-quality training courses, which would be chosen without financial support. For example: the number of high-quality training courses offered in base scenario was 409, which means, that ca. 85% of the training courses offered after subsidy would be offered anyway (but at an effectively higher price). This is not necessarily negative, because the money saved on ILAs can be used for funding training in the future. Though one might anticipate that training providers who expect that they would have been chosen anyway, may increase their offer price by €150 to benefit from the ILA funds.

4. Mandatory employer co-financing

As might be expected, the increase in employer funding also increased the number of individuals receiving training from 94% to 97.5%. The number of individuals who were not allocated any training course fell by more than half (from 36 to 15).

In line with the increase in employer funding, the total financial cost of all allocated training courses increased from 309 822€ to 373,890€. Interestingly, however, the overall increase in funding did not affect the increase in the total number of courses allocated. On the contrary, it decreased from 1 741 to 1 726, i.e. by less than 1%. In addition, training costs per person per most recommended training course increased from €203 to €272.70, suggesting that AI was not allocating a higher number of training courses to individuals, but training courses at a higher cost. If we assume that a higher course cost also reflects a higher course quality, we can conclude that the increase in funding has contributed to an overall increase in the quality of the training courses allocated.

Looking at the number of recommended training courses broken down by framework combinations, we can see that the combination of measures, such as an increase in funding from the employer and taking into account the individual learning goals, only resulted in an increase in the number of training courses allocated for the development of language and digital skills (CEFR and DigComp). Conversely, there was a decrease in the number of training courses for the development of financial literacy and entrepreneurial skills (FinComp and EntreComp). This trend is also confirmed by the results showing changes in competence levels. The number of level changes, which increased by more than 30% in the CEFR and almost 19% in the DigComp frameworks, indicates that employees whose employers contributed to their training were particularly interested in investing these financial resources in the development of the transversal skills that are currently among the most in demand in the labour market.

In the case of indicator 6, the implementation of the professional concept resulted in a significant increase in interest in language skills development (CEFR) and a significant decrease in attending courses related to FinComp in all target groups (general sample, NEETs, Low-qualified).

Overall, the number of NEETs who were allocated training courses in various groups of skills was significantly lower after we implemented the professional concept, suggesting that the increase in funding from employers had a negative impact on this particular target group.

V.5 Conclusions for the D-ILA data model

Once implemented in practice, the results of the simulations will heavily depend on the quality of the data and its correspondence with reality.

We modelled the synthetic data on real personas and considered the socio-demographic specificity of the Visegrad countries. Additionally, we tested for differences between the sub-populations. Having said that, it remains beyond question that adjustments to the data used are in all cases necessary. For example: the data in the model allows introducing any value of state or individual support and these values vary significantly. At the same time, country A may decide to provide each citizen with an equal amount of entitlement while country B may decide to differentiate the support based on some criteria (e.g. assigned target group).

The testing and development have yielded in the following conclusions for the D-ILA data model:

- The D-ILA data model can only modify certain aspects once the algorithm is trained. For example, changing the preference for learning goals requires pre-determined constraints.
- The indicator system developed is sensitive to changes, which can lead to false conclusions, such as the example where a drop in training offers affected all other indicators.
- The model cannot fully capture real-world complexities, such as individuals' resistance to mandatory contributions for education, and requires labour-market experts to interpret results accurately.
- A limited number of training courses in the sample did not lead to a proportional decrease in assigned individuals, indicating an unexpected robustness of the algorithm. This may suggest that limitations concerning the quality of the matching proposed would be required.
- Up to five different 'framework category' entries per course can lead to unclear results and ambiguous statistics. At this phase of work a limit to the number of categories was needed, yet it was limited to soft skills and related frameworks. The actual ILA data model will require considering specific skills and allowing multiple skills and/or categories to be used (e.g. ESCO).
- The model includes only economically active individuals, suggesting a need to consider whether the group of inactive individuals should be analysed.
- The training database assumes the reliability of information regarding quality assurance mechanisms. This should be the topic of further considerations when applying the data model.
- The AI model's recommendations are significantly influenced by the indicated training needs, demonstrating the model's operational effectiveness.
- The algorithm maintains a similar rate of assigning training courses to individuals even when the number of courses is reduced. If the accuracy of these matches is low, the possible solution is to include more examples in which there are 'no matches' in the AI training data.
- Future use of the model should consider adjusting financial support for unmotivated groups to understand their impact on motivation and participation in education.
- To better analyse the effects of the AI run and avoid false conclusions, new indicators should be developed to assess changes more accurately.
- A larger and more differentiated list of training courses should be used to retrain the AI.
- Users need to map their population (sample) in detail to make relevant conclusions and avoid misleading evaluations.
- Future attempts should review how scenario options are modelled, considering potential limitations and hidden influences on recommendations.
- The size of the sample of learners (600) seemed to be sufficient for the task at hand. Though for prognosing which may have real impact and in cases with much larger training databases, the sample needs to be much larger.

Clearly, the model developed is open to many modifications and improvements. However, even at this early stage of development, it has proven to be a useful and insightful tool. The size of the sample and number of parameters in this model have provided limitations to the types of AI that can be used. Although even statistical reasoning may provide good results.

Funded by
the European Union

VI. Methodological guide for the application of the model

The D-ILA data model and executable programme developed within the project can be applied in a different context. This part shows how to do it and provides a step-by-step guide to using it. Although we tried to describe the steps in a clear manner and provide help for D-ILA model users, one must keep in mind that applying the model to one's own needs will require a serious investment of time. The D-ILA project is, in part, about making that investment smaller than it would otherwise be: instead of starting from scratch, you receive a 'template' for a data model and proof that it can work.

Before attempting to modify the data structure and/or datasets, we propose using the default data and executable programme provided. The best way to understand the logic of the D-ILA data model is to use it.

Please keep in mind that, if you make deep modifications to the data structure, you will need to retrain the AI - which will require involving a data scientist to run a new algorithm for you.

VI.1 Overview

The D-ILA project team proceeded as follows:

One table contains the names and characteristics of the data fields selected by the experts. The data fields are part of the D-ILA data model describing adult education. According to the structure defined here, the data series edited by experts are produced. The table not only contains the data fields and their characteristics, but also defines precisely the values that each field can take. So, when the editing data series, experts can choose from the values specified here. Of course, some of the data fields must be free to complete, such as the cost of training data field.

The other table contains data fields and their characteristics that, according to the experts, are suitable for describing the characteristics of individuals participating in adult education. Similarly to the table of adult training courses, the value sets of each data field are included in this case, which determine what values experts can choose from when editing the data series. The project team strived to create a data model and predefined values of the data fields so that they resembled the parameters of real training courses.
The last tab of the table ("skipped") lists data fields that were not included in the data model during expert discussions but may be relevant for other users.

When the data table describing the characteristics of the training and the data table describing the characteristics of the training participants were completed, the preparation for the use of AI followed. The AI learned from the groupings of individuals and training courses created manually by experts, based on which rules the groupings should be made. What the AI learned was applied when individuals and courses were in separate tables and the task was to match them.

VI.2 Sequence of steps to make your own D-ILA data model

In this chapter, we present the sequence of steps by which the D-ILA data model can be built. During the rebuilding of the data model, it is of course advisable to use data fields, data field value sets and

Funded by
the European Union

data rows suitable for your purpose - while the sequence of steps also follows the sequence described here during the construction of a new model.

1. **Learning about D-ILA data fields**

Let's get to know the data fields that the Partnership's experts have incorporated into the model. There are basically two types of data groups. One is the data fields for describing the characteristics of the individuals participating in the training, while the other is for the characteristics of the adult training. You already read about them in the previous chapters of this guide, but now it's time to browse the files with actual data. We recommend reading some data rows and decoding the value sets.

2. **Defining own needs and expectations of the model, assessing data availability**

Feel free to modify the data fields according to your own needs: delete those that are not needed, and insert new data fields as needed. In this phase, it is worth checking the availability of the data. Of course, it is not advisable to use data fields in the model that you will not be able to fill with real data later.

3. **Designing value sets**

When we are done compiling the data fields describing the characteristics of individuals and adult training courses, the value sets of each data field are determined. The value set determines what values the given data field can take when the model is filled with real data. According to our experience, the definition of the value set of the data fields is an extremely important task, it reflects fundamental political and, in many cases, even social political aspects. In the following part, we present some examples of dilemmas concerning designing the value sets.

4. **Development of a commensurability matrix**

One of the most important aspects in developing the data fields and value sets is testing. Do this by creating individual test entries for both training opportunities/qualifications and data on individuals. This will reveal difficulties in applying your value sets or gathering some data. In designing the data model (data fields and value sets) a design thinking approach, e.g. thinking of real-life applications, may be very useful.

At this stage you may also think of the links between datasets: which fields can be used to find a match between a training course and an individual (e.g. individual's entitlement and training cost). In our project, we developed a commensurability matrix (see Annex), which helped us see the relation between these two datasets.

5. **Gathering or creating training data sets**

The finalisation of the data fields and their value sets is followed by the production of the data sets required for training the AI. First, let's prepare the data sets describing adult education. This is done by entering the data that the data fields allow, and the value sets defined for each training course. If, for example, the properties of a training course can be described with 30 types of data fields, then a maximum of 30 data (thinking in a table: cell values) must be entered for a given training course. Let's make an effort to fill the table containing the data lines of the training courses with the real data of the real training courses available on the market (as previously indicated, the D-ILA data model contains realistic, but not real, training data). In case of using available data, remember to see how many empty or erratic values it contains: you may decide to leave it as it is, correct or delete from the dataset (depending on the type of these errors / peculiarities).

Funded by
the European Union

6. Gathering or creating individual data sets

Let's carry out this step for individuals. This creates another data table, which contains the data (cell values) describing the characteristics of the individuals. Each data line contains the data of one individual.

7. Matching the data on individuals and training

The two data tables are connected in this step. Linking means that we assign at least 1 training course to each individual. It is possible to assign multiple training opportunities, but it is also possible to assign no training in some cases (The technical implementation of the assignment can be found in the sample files). The matching of individuals and courses proved to be the most difficult task during the implementation of the project. The difficulty of matching is reduced if we work with data from real individuals and real training courses. In this case, it is advisable to create pairs where we already know it was a success. After all, we want to teach AI what we know works. Simply put, we show the AI what are the good practices that the AI should follow when choosing a training course for a given individual.

8. Training the AI on the prepared data.

This sets the stage for running the AI to learn the patterns shown by the good practice data. Keep in mind that the training of the AI will require more data if more parameters are to be taken into account, i.e. because the patterns of matching become more complex. In order to see if the AI provides sensible results, you need to have more data prepared in advance (the training and test parts of the dataset) or a lot of time to assess the results of the AI training.

In the future, we will use the data table describing the trained AI and the characteristics of the training. It is also necessary to compile a data table that contains data describing the characteristics of the individuals for whom we expect the AI to define the proposed training.

This step is therefore a technical step. This is where the data tables edited by experts in the previous point - or in other cases based on real data - are "shown" to the AI so that it can analyse and learn the relationships that can be read from the tables. In fact, this is where the advantage and meaning of AI is revealed, because it can also discover and apply such relationships during subsequent runs that human resources are only able to do to a limited extent or not at all.

9. Preparing the modelling dataset for the different scenarios

After the AI is trained, you will need the data to run the model and see what the changes introduced in a given scenario will result. For that purpose, it is necessary to use the same dataset used for AI training. You may use a set containing data on individuals which reflects the real population and/or it is also possible to create such a dataset using data augmentation techniques - the point is to have a dataset that will resemble the initial individuals' dataset in terms of distribution of relevant variables. In simple words, if the AI was trained on one population, it may not provide correct matching for other, drastically different populations.

Because of the sample size limitations in the project, we decided not to create a new training data set, but use the original one. This creates relevant restrictions. **Limitations to the approach taken.** Following from the applied methodology, there are strong restrictions here. The trained AI can only give suggestions if the data table describing the characteristics of the individuals contains the same data fields and data field value sets as the data table with which the AI was trained in step 5.

Furthermore, the training data sheet must be the same as the one we used before. So, in the case of training, not only the data fields and value sets must match, but also the data rows themselves.

10. Start modelling policy interventions

In the last step, we run the trained AI and generate training suggestions for each individual listed in the table created in step 8. (The AI-based solution used in the D-ILA data model performs learning and matching on a model resulting from the first three steps)

VI.3 Products that help to implement the sequence of steps

In order to reproduce the sequence of steps presented in the previous chapter more easily, we will present the resulting products that we ourselves created within the framework of the project. When building and testing the D-ILA data model for your own purposes, we recommend creating similar products - using them as samples.

1. **Presentation of the data field structure of the D-ILA data model and the possible values of each data field – adult training courses**

The table contains the names and characteristics of the data fields selected by our experts. The data fields selected by the experts are part of the D-ILA data model describing adult education. According to the structure defined here, data series edited by experts are produced. The table not only contains the data fields and their characteristics, but also defines precisely the values that each field can take. So, when editing data series, experts can choose from the values specified here. Of course, some of the data fields must be free to complete, such as the cost of training data field.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5274997

2. **Presentation of the data field structure of the D-ILA data model and the possible values of each data field – adult learners**

The table contains the data fields and their characteristics that – according to our experts – are suitable for describing the characteristics of individuals participating in adult education. Similarly to the table of adult training courses, the value sets of each data field are included in this case, which determine what values experts can choose from when editing data series. The last tab of the table ("skipped") lists data fields that were not included in the data model during expert discussions, but may be relevant for other users.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5275009

3. **Data series of the D-ILA data model edited by experts – adult education**

According to the original idea, our team of experts was commissioned to define 100 training courses. Finally, the data of 106 training courses was edited. The reason for this is that during the analysis of the training offer, gaps were revealed that were not covered by the 100 training courses, which is why new training courses were included in the table. It is important to note that, according to the concept of the project, the training courses in the table – described in the individual data series – is training imagined by experts, so it is not possible to enrol in such training courses in practice. On the other hand, the data model and the predefined values of the data fields ensure that real training courses can have the same parameters. Of course, the professional experience of our team of experts also served the purpose of including realistic training courses in the table containing the edited data.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5275092

Funded by
the European Union

4. **Data series of the D-ILA data model edited by experts – 900 individuals participating in adult training**

The table contains the data of 900 trainees envisioned and edited by experts. Each data series is defined by using the data fields and data field value sets presented in the D-ILA data model. Our experts filled in each data individually to the best of their professional knowledge and experience. This ensures that the data characterises realistic individuals. One of the most difficult tasks – needing the most creativity and human resource – of the project was to construct this table.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5275175

5. **Data tables for training and testing AI**

In accordance with the project plan, the data sets for the description of the 900 trainees produced by our team of experts were divided into two groups. 300 data sets will be used to train the AI while 600 data series will be used to test the trained AI. In this case, the selection was made using the simplest possible method: each row was assigned an ordinal number using a random number generator, and then, putting the data rows in ascending order based on the random sequence number, we formed a table for training AI from the first 300 data rows. The table accessible via the link already contains the 300 and 600 data series in separate tabs.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5275111

6. **Statistics from the 106 adult training course data tables – interactive data visualisation (Microsoft PowerBI)**

The experts of the project edited the data of 106 training courses. Based on this data, an interactive statistical data visualisation was created, which requires the free desktop version of Microsoft PowerBI to be installed on your computer. The goal of data visualisation is to make data validation easier. With the help of the presented statistics, external experts who are not familiar with the project can easily and quickly get a comprehensive picture of the training offer consisting of edited training courses. It is important to emphasise that following the objectives of the project, the training offer was compiled by our experts in the most realistic form possible.

https://drive.kifu.hu/index.php/s/ej4YtKPAxJZQXkp?path=%2FD-ILA (D-ILA trainings stat_v3.pbix)

7. **Statistical data of the 106 adult training data tables – traditional .pdf statement**

The following .pdf is designed for those who do not use Microsoft's PowerBI software. The .pdf file contains a preconfigured, static state of the interactive interface.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5223956

8. **Statistical data of 300 adult learners for AI training – interactive data visualisation (Microsoft PowerBI)**

Microsoft PowerBI desktop is required to open the data visualisation. The project experts constructed the data of a total of 900 imaginary individuals. Of these, 300 have been selected to whom experts in WP4 manually assigned a course out of the 106 available training courses. The 300 individuals and their assigned training courses will be used to train AI. The data visualisation was created to provide an easy and quick overview of the statistical characteristics that allow external experts and interested parties to make sure that the sample created by editing describes a truly realistic population. So AI is trained using data that is realistic.

Funded by
the European Union

https://drive.kifu.hu/index.php/s/ej4YtKPAxJZQXkp?path=%2FD-ILA (D-ILA participants stat_v3.pbix)

9. Statistical data of 300 adult learners for AI training – traditional .pdf statement

The following .pdf is designed for those who do not use Microsoft's PowerBI software. The .pdf file contains a preconfigured, static state of the interactive interface.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/KgmWBajwAqkWtcq?fileId=5223949

10. Statistical data of 600 adult learners to test the trained AI – interactive data visualisation (Microsoft PowerBI)

Microsoft PowerBI desktop is required to open the data visualisation. This data visualisation shows edited data of imaginary individuals that was used to test the trained AI in WP4. The aim of data visualisation is to enable external experts and interested parties to easily and quickly ascertain how realistic the population is based on statistical characteristics.

https://drive.kifu.hu/index.php/s/ej4YtKPAxJZQXkp?path=%2FD-ILA (D-ILA participants stat_v3_600 .pbix)

11. Statistical data of 600 adult learners to test the trained AI – traditional .pdf report

The following .pdf is designed for those who do not use Microsoft's PoweBI software. The .pdf file contains a preconfigured, static state of the interactive interface.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5280767

12. Interpreting the D-ILA Data Model – presentation

The slides present the main features of the ILAs, the operation of the D-ILA data model, and how AI-based solutions are applied. Subsequently, the statistical characteristics of the data series generated by experts of trainees and training courses are presented.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5274783

13. Table for editing data of training courses for the development of transversal competences

You can use the table to test how to determine a training course by providing descriptive data. Code tables help fill it in, so you can choose from a drop-down list for each data field. The project experts used a similar data editing interface to construct data from a total of 106 imaginary courses.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5280806

14. Table for editing data of participants in training

You can use the table to test how to define an individual participating in a training course by providing descriptive data. Code tables help fill it in, so you can choose from a drop-down list for each data field. The project's experts used a similar data editing interface to construct the data of a total of 900 imaginary individuals.

https://drive.kifu.hu/index.php/apps/onlyoffice/s/ej4YtKPAxJZQXkp?fileId=5280815

15. Individuals participating in adult education – data generated by algorithms

Funded by
the European Union

Within the framework of the project, the expert team is primarily responsible for compiling the data series of individuals participating in adult education. Nevertheless, we also tried to generate data. The peculiarity of the resulting data series is that the generation was based on a predefined data distribution, i.e. contrary to the data series approach used by experts, in this case the generation took place at the level of data fields. The document contains both the source code of the algorithm used for generation as well as the statistical characteristics of the generated sample compared with the statistical characteristics of the data edited by experts.

https://drive.kifu.hu/index.php/s/ej4YtKPAxJZQXkp?path=%2FD-ILA
(participant_data_generation.html)

16. Adult training courses – data generated by algorithms

In the case of adult training courses, we experimentally tried the production of data series using an algorithm. In this case, too, the distribution of a data field is the starting point. The document contains, both the source code of the algorithm used for generation as well as the statistical characteristics of the generated sample compared with the statistical characteristics of the data edited by experts.

https://drive.kifu.hu/index.php/s/ej4YtKPAxJZQXkp?path=%2FD-ILA
(training_data_generator_plots.html)

VI.4 Using the D-ILA data model for the evaluation of policy intervention scenarios

The D-ILA data model allows the introduction of changes to the data used for modelling and, through those changes, the evaluation of various policy scenarios.

There are two datasets in use:

- The AI-training dataset, consisting of two subsets: the individual data (300) and training course data (106)
- Modelling (testing) dataset, consisting of the individual data (600).

In most cases we make changes only in the modelling dataset. Here we can manipulate all the parameters related to the ILAs: the amount of entitlements in an individuals' account as well as the learning goals (content/topic preferences in learning), time available, preferred mode of learning etc.

The limitations of the D-ILA model are such that changes concerning the training dataset are not modelled properly, i.e. the model is not fit for modelling this type of intervention. One can expect that cases in which 10% of the training courses are deleted from the dataset will result in reasonable suggestions, while substituting 50% of the training courses with new ones (of different characteristics) would not. Modelling some of such scenarios would require (again) training the model, but on a much wider group of training offers. However, some scenarios can be modelled through a change in the individuals' dataset, for example a decrease in the average relative cost of training can be modelled by manipulating the amount of the training entitlements (i.e. an increase in individuals 'budget' at fixed prices).

We propose to follow the steps described below in order to keep track of the policy scenarios, their results and their evaluation. We propose to use a simple template for each simulation, which can later be retrieved. You can use our template for it or modify it (see Annex I: AI-runs reports).

1. Write down the rationale, logic and expected results of the intervention. Name the scenario, so that it will be easier to retrieve later on, when many scenarios have been analysed.
2. Decide on the necessary changes in the AI training dataset and the modelling dataset. Describe them briefly.
3. Take a moment to reflect on the indicators which will be relevant for the evaluation of the results, and if the 'budget allocation parameter' is going to be used.
4. Make the changes in the modelling dataset.
5. Save the changed tables in the folder 'recommender' in the file titled 'MASTER database for matching_v5_OK'.
6. Run the AI and generate the results.
7. Compare the results of the generated simulation with the baseline scenario to evaluate the results of a given policy intervention scenario.
    1. Use the indicators proposed and put them in the AI-run report.
    2. If you need additional indicators, you can calculate them in the .xls output files (results of the baseline scenario and the policy intervention scenarios are saved in .xls format for easier access and calculations)
8. Evaluate the results.

VI.5 Q&A for the D-ILA data model

What is the relationship between the data used in the model and reality?

One of the important experiences gained from the project is that, with the help of the data, we can only approximate the processes taking place in practice. The 42 in-depth interviews carried out in the framework phase of the project highlighted that, especially at the company level, the pairing of individuals and training courses is the result of human interactions that are difficult to describe with data. This does not mean that the data cannot be used at all, it just means that you need to be aware of the limitations of the data-based approach and its potential.

How realistic is the data used in the D-ILA data model?

The D-ILA data model contains data of non-real individuals and non-real training courses. There are many reasons for this, but perhaps the most important is that - for the time being - no real data exists regarding transversal competences. We tried to generate some but, according to our preliminary expectations, not very successfully. So experts created edited data. The validation was horizontal and vertical. On the one hand, we looked at an individual or training course during the validation stage, presenting an arbitrary set of data to anyone, they all said that yes, there can be a training course that can be described with such data. The distribution of data fields is more debatable,

What do we mean by realistic data?

The D-ILA data model uses realistic data. This means that anyone who is shown the data describing the attributes of a given training course in the D-ILA data model can say, "Yes, although I do not currently know of such a training course, I can fully accept that such a course could exist." In another light: if we have 10 training courses, of which 9 are real and 1 is realistic, based on the data describing the course, not even a trained expert can tell which is the realistic training data set.

We used realistic data in the D-ILA data model because we could not find available data on real training courses that develop transversal competences. This only seems to undermine our model. Let's reverse

Funded by
the European Union

the logic: with the help of realistic data, we can define a training need, which we can then advertise on the training market. The training developers can respond to this and develop training tailored to the needs. Based on the 42 in-depth interviews conducted as part of the project, this approach is unfortunately missing from the adult education market in many cases. "There is real training here, it's true that it's not exactly what we'd expect, but it's better than no training at all."

The D-ILA data model does not include training organisation data. What is the reason for this?

The properties of the courses included in the D-ILA data model are described with realistic data. During the implementation of the project, we came to the conclusion that the low number of data rows does not allow the inclusion of training organisation data fields in the model, because, for example, specifying the starting time and place of a training course would drastically limit the number of individuals who can be paired with the training course, thus the model would be inoperable. The training defined in the framework of the project (107 units) and their assignment to individuals means, in practice, that the training recommended by the D-ILA data model is considered a type of training, which provides a search condition for the training courses available to the given individual with a specific start date and location.

Is the D-ILA data model training suitable for recommendation or is its primary purpose to support data-based decision-making and modelling?

It is suitable on a small scale, i.e. at the level of a specific company. It is trained with real data, where the training has worked before, and thus the system provides a training recommendation that is already working well for a new employee. At the same time, we use the D-ILA data model in the project to analyse policy decisions, that is, we change the input data and examine how the output changes as a result. The changes are analysed with the help of statistical evaluation tools. We recommend viewing the AI run logs.

How to decide on the value sets for various data fields?

The value sets play an important role in designing the data model. They may limit or enable various use cases of the data. Usually, there is a tension in the comparability and accuracy of a given value set. Using an open value set (e.g. allowing any text input in a given field) may be a very good decision for the training name, but a poor one for expressing a competence level: even if you have an accurate description of an individuals' competence level, you may not be able to process and use this information later on. It is good to think about the granularity of the data, for example you may want to express a given characteristic: in a 1-5 scale or a 1-10 scale or you may want to use different levels of a hierarchical classification or a taxonomy, such as the ISCO or national occupations classifications. In both cases, the decision should depend on the use cases. If possible, refer to existing typologies or classifications, however, keep in mind that some of them may not be suitable for your purpose (e.g. classification of training in terms of NACE classification may turn out to be meaningless).

How much data will I need to train the AI?

The amount of data will depend on the quantity of data fields and their value sets, as well as on the quality of the data itself. This question needs to be discussed with an AI engineer as early as possible - because various methods and approaches can be used, the actual amount of data can vary.

Possibly not all patterns require an AI algorithm to perform filtering of AI-based results. For example, you may use AI for matching individual characteristics such as education, vocation or learning priorities with training contents/titles, but you probably can use a simple algorithm to then filter the results

Funded by
the European Union

based on the comparison of the training price and the training entitlement on one's individual account. It is also worth noting that to train the AI you need to be able to see if the training worked: for that you would usually randomly divide the dataset into two parts prior to training: the training and validation/testing subsets. After training the AI, you can run it on the testing subset and compare the results with your data - this is a standard practice and the rule of thumb is 80/20.

VII. Use cases for the D-ILA data model

Use Case 1: Personalized Learning Pathways for Upskilling Workers

In this scenario, the D-ILA data model is applied to help employees in industries facing rapid technological changes (e.g., automation and digitalization) identify relevant upskilling opportunities.

Context: A government agency or private employer wants to provide support to employees at risk of job displacement due to technological advancements. The goal is to use ILAs to fund training programmes that will help these employees acquire new digital skills and remain competitive in the job market.

How the D-ILA Model Works:

1. Data Collection: The model gathers data on the employees' current skills, job roles, educational backgrounds, and preferred learning methods. Information about the available training programmes, including the content, cost, delivery mode, and duration, is also inputted into the system.
2. AI Matching: AI uses this data to match employees with training programmes that are most aligned with their skill gaps and career goals. For example, an employee who lacks skills in data analysis but shows proficiency in other digital skills may be matched with a course on advanced data analytics.
3. Outcome: The model provides personalised training recommendations that meet both the employee's learning needs and the company's strategic goals (e.g., improving digital literacy). The employees are guided toward relevant courses that align with the evolving demands of their industry, ensuring they stay employable despite technological changes.

Use Case 2: Supporting Lifelong Learning for Disadvantaged Groups

This scenario focuses on using the D-ILA model to promote lifelong learning for disadvantaged groups, such as unemployed individuals, low-income workers or migrants, ensuring they have access to education and training.

Context: A government or non-profit organisation seeks to increase participation in adult education for disadvantaged groups. By providing targeted financial support through ILAs, the organisation wants to help these individuals develop essential skills for employment.

How the D-ILA Model Works:

1. Data Collection: The model collects data on the learners' demographic profiles, including age, income level, employment status, and prior education. It also gathers information about the training programmes available, such as vocational skills, language courses, and soft skills training.

Funded by
the European Union

2. AI Matching: The AI algorithm identifies courses that best suit the needs of the learners, considering factors such as geographic location, language proficiency, and the level of support required. For instance, a migrant worker with limited language skills may be matched with a local language course followed by a vocational skills programme.
3. Outcome: The D-ILA model ensures that the most disadvantaged learners receive tailored learning recommendations that help them improve their employability. The model also accounts for accessibility, ensuring that learners with specific needs, such as remote or part-time learning, receive appropriate course recommendations.

Use Case 3: Policy Evaluation for Government Education Funding

In this scenario, a government uses the D-ILA model to evaluate different education policy interventions and optimise funding allocation for adult learning.

Context: The government is interested in determining the most effective way to allocate funding for adult learning programmes. They want to understand the impact of various policy decisions, such as increasing funding for specific skill areas (e.g., green skills or ICT skills) or targeting specific groups (e.g., unemployed adults).

How the D-ILA Model Works:

1. Data Collection: The model collects data on current adult learners, including demographic information, employment status, and educational backgrounds, alongside available training programmes and their costs. The model also integrates policy variables, such as funding levels and eligibility criteria for ILAs.
2. AI Simulation: AI runs simulations based on different policy interventions. For example, the government could test how increasing funding for green skills training would affect enrolment and outcomes compared to focusing on digital skills. The AI evaluates the possible outcomes of each scenario, such as the number of learners reached, the success rate of the programmes, and the impact on employment.
3. Outcome: The government receives data-driven insights into how different funding levels and policy changes would influence adult learning outcomes. This allows policymakers to make informed decisions, ensuring that public funds are allocated in the most efficient way to maximize learner participation and skill development.

Use Case 4: Corporate Training and Workforce Development

In this scenario, a large corporation uses the D-ILA data model to provide personalised training opportunities for employees as part of their internal workforce development initiatives.

Context: A corporation wants to boost the skills of its workforce by offering targeted training programmes. The company aims to enhance employee engagement, retain top talent, and ensure that its workforce is equipped with the latest skills to meet evolving business needs.

How the D-ILA Model Works:

1. Data Collection: The model collects data on employee skills, job roles, performance reviews, and career aspirations. Training programmes offered by the company—ranging from leadership development to technical certifications—are also integrated into the system.

Funded by
the European Union

2. AI Matching: AI matches employees with the most relevant training programmes based on their current skill levels, job performance, and career goals. For instance, a mid-level manager aspiring to move into a leadership role might be recommended a combination of leadership training and advanced project management courses.
3. Outcome: The company can provide highly personalised training plans for each employee, ensuring they receive the right skills for career progression. This approach increases employee satisfaction, improves performance, and helps the organization

Use Case 5: Regional Skills Development for Local Economic Growth

In this scenario, a regional government or development agency uses the D-ILA model to focus on workforce development that aligns with the economic needs of a specific region.

Context: A regional development agency wants to stimulate local economic growth by ensuring that its workforce is skilled in industries that are critical for the area, such as green energy, agriculture, or advanced manufacturing. The goal is to ensure that local residents receive training that directly supports the region's key industries and economic priorities.

How the D-ILA Model Works:

1. Data Collection: The model collects data on the region's labour market needs, available training programmes, and the skills of the local population. For example, the region might have a demand for workers skilled in renewable energy technologies, but the local workforce lacks the necessary qualifications.
2. AI Matching: AI identifies training programmes that align with both the regional economic priorities and the skill profiles of local residents. The model helps match individuals with training opportunities in industries that are projected to grow in the region. For instance, individuals with basic technical skills might be recommended advanced courses in solar panel installation or wind turbine maintenance.
3. Outcome: The region can efficiently allocate resources to ensure that training programmes are well-aligned with labour market needs. By using the D-ILA model, the regional government can foster a highly skilled workforce that supports local industries, attracting new investments and driving economic growth.

Use Case 6: Educational Institutions Optimising Programme Offerings

In this scenario, an educational institution, such as a university or vocational training centre, uses the D-ILA data model to assess and optimise its programme offerings based on learner demand and labour market trends.

Context: A university or vocational training centre wants to ensure that its course offerings are responsive to student demand and aligned with the current labour market needs. The institution seeks to improve enrolment in its programmes and ensure that graduates are well-prepared for high-demand job sectors.

How the D-ILA Model Works:

1. Data Collection: The model integrates data on student interests, skill levels, and career goals, as well as regional and national labour market trends. It also incorporates data on the institution's available programmes, including course content, duration, and outcomes.

2. AI Matching: AI analyses this data to identify gaps between student demand and the institution's offerings. For example, if there is a growing demand among students for courses in data science and cybersecurity, but the institution only offers a few basic IT courses, the model will highlight the need to expand these programmes.
3. Outcome: The institution can adjust its course offerings to better match the needs of students and the labour market. This data-driven approach helps the institution stay competitive, increase enrolment, and ensure that its graduates are equipped with the skills required by employers.

Use Case 7: Personalised Career Pathways for Youth

This scenario focuses on using the D-ILA model to help young individuals, particularly school leavers and recent graduates, plan personalised career pathways through relevant education and training.

Context: A government body or educational institution aims to support young people in transitioning from education to employment by providing guidance on career pathways and appropriate training options. This is especially important for young people who may not have clear career goals and need support in identifying relevant opportunities.

How the D-ILA Model Works:

1. Data Collection: The model gathers data on the learners' academic backgrounds, career aspirations, interests, and available training programmes. It also incorporates labour market data to ensure that the career pathways align with current job opportunities.
2. AI Matching: AI analyses the data to suggest personalised career pathways for each learner. For instance, a student with an interest in technology but limited technical experience might be recommended foundational IT courses, followed by internships or certifications in coding or cybersecurity.
3. Outcome: The learners are provided with a step-by-step career roadmap, detailing which courses and certifications they should pursue, how much funding is available through their ILAs, and the potential job outcomes. This personalised guidance helps the learners transition smoothly from school to work, ensuring they are on a path that leads to meaningful employment.

Use Case 8: Lifelong Learning for Older Workers

In this scenario, a national or regional government uses the D-ILA model to encourage lifelong learning among older workers, helping them remain competitive in the job market and adapt to changing industry demands.

Context: As populations age, many older workers need support to remain employed and adapt to technological advancements. The government aims to provide financial support for training programmes through ILAs to help older workers upgrade their skills, particularly in sectors like digital technologies and management.

How the D-ILA Model Works:

1. Data Collection: The model collects data on the skills, experience, and career histories of older workers, alongside available training programmes that focus on skill upgrading or reskilling for workers aged 50 and above.

Funded by
the European Union

2. AI Matching: AI matches older workers with programmes that address both current job demands and future trends. For example, a 55-year-old worker in manufacturing could be matched with a programme on managing automated systems or retraining in areas like logistics and supply chain management.
3. Outcome: Older workers are provided with personalised training recommendations that allow them to remain employable in industries where technological advancements are changing job roles. This helps address age-related unemployment and ensures that experienced workers can continue contributing to the economy.

Use Case 9: Skills Matching for Migrant Integration

In this scenario, the D-ILA model is used by a government agency or Non-Governmental Organisation (NGO) to help migrants integrate into the labour market by matching them with training programmes that build on their existing skills and prepare them for local job opportunities.

Context: Migrants often face challenges in finding employment that matches their qualifications due to language barriers, lack of recognition of foreign credentials, or gaps in local knowledge. By providing ILAs, the government or NGOs can help migrants access relevant training programmes that improve their chances of securing employment in their new country.

How the D-ILA Model Works:

1. Data Collection: The model collects data on migrants' education levels, professional experience, language skills, and career goals. Information on available training programmes, such as language courses, vocational training, and professional certifications, is also incorporated.
2. AI Matching: AI analyses this data to recommend training programmes that will help migrants transition into the local job market. For example, a migrant engineer might be matched with a local certification programme to meet regulatory standards or language courses to improve employability.
3. Outcome: Migrants are given tailored learning pathways that consider their existing skills and local labour market needs. This approach accelerates their integration into society, enhances their employability, and reduces the skill gaps in the local economy.

Use Case 10: Rapid Response to Economic Shocks

In this scenario, a government uses the D-ILA model to respond to economic disruptions, such as the loss of jobs in a specific sector, by quickly identifying training opportunities for displaced workers.

Context: During economic downturns or sector-specific disruptions (such as the decline of coal mining or the automation of manufacturing jobs), many workers find themselves unemployed or at risk of losing their jobs. The government seeks to rapidly retrain these workers to help them transition into new roles in growing industries.

How the D-ILA Model Works:

1. Data Collection: The model gathers data on displaced workers, including their skills, job history, and willingness to transition into new industries. It also incorporates data on growing industries and training programmes relevant to those industries.
2. AI Matching: AI identifies the most appropriate training programmes for each displaced worker based on their transferable skills and emerging job opportunities.

For example, a manufacturing worker may be guided toward retraining in logistics, healthcare, or renewable energy.

3. Outcome: The D-ILA model provides rapid, data-driven support to displaced workers, helping them re-enter the workforce with new skills suited to growing industries. This minimises long-term unemployment and helps stabilise the local economy during times of economic disruption.

Use Case 11: Optimising Corporate Training Budgets

In this scenario, a large organisation uses the D-ILA data model to optimise its internal training budget, ensuring that resources are allocated to the most impactful training programmes for employee development.

Context: A company wants to maximise the efficiency of its corporate training budget by ensuring that its investment in employee training aligns with both employee development needs and the company's strategic goals. The company also seeks to ensure that its employees remain competitive and innovative by acquiring the latest skills.

How the D-ILA Model Works:

1. Data Collection: The model gathers data on employee skills, training history, performance evaluations, and development goals. It also includes details on the company's available training programmes and their associated costs.
2. AI Matching: AI analyses the data to recommend the most cost-effective and impactful training programmes for each employee. For instance, an employee in the sales department might be recommended courses in digital marketing, negotiation, or leadership, depending on their career goals and the company's needs.
3. Outcome: The company optimises its training budget by prioritising training programmes that deliver the highest value for both the employee and the organisation. This ensures that employees receive the most relevant training, leading to better performance, retention, and overall productivity.

Use Case 12: Supporting Entrepreneurship Through Tailored Training

In this scenario, the D-ILA model is used by a government or NGO to provide aspiring entrepreneurs with personalised training that helps them develop the skills necessary to start and grow their own businesses.

Context: A government or NGO wants to support entrepreneurship by helping individuals who aspire to start their own businesses gain the skills they need. This could include training in business planning, financial management, digital marketing, and other essential entrepreneurial skills. The goal is to empower individuals to create successful businesses, particularly in underserved regions.

How the D-ILA Model Works:

1. Data Collection: The model collects data on individuals interested in entrepreneurship, including their current skill levels, business ideas, and industry focus. It also gathers information on available entrepreneurship training programmes, including those that focus on finance, management, and legal requirements.
2. AI Matching: AI analyses the data to provide personalised recommendations for each aspiring entrepreneur. For instance, someone with a background in creative industries

might be recommended training on digital marketing and e-commerce, while someone with a technical background might be directed toward training in business finance and operations management.

3. Outcome: Aspiring entrepreneurs are provided with a personalised learning pathway that equips them with the skills necessary to launch and grow successful businesses. This helps foster entrepreneurship in the community, driving innovation and economic growth, especially in regions that lack established support systems for new businesses.

Use Case 13: Flexible Learning Pathways for Part-Time Workers

In this scenario, the D-ILA model is used by a government or educational institution to help part-time workers access flexible training programmes that align with their work schedules.

Context: Many part-time workers have limited access to training opportunities due to time constraints and irregular work hours. A government or educational institution seeks to provide part-time workers with access to flexible learning opportunities that fit into their schedules, helping them upskill or transition into full-time employment.

How the D-ILA Model Works:

1. Data Collection: The model collects data on part-time workers' employment schedules, current skills, and career aspirations. It also gathers information on flexible training programmes that offer online or evening courses, short-term certifications, or modular learning options.
2. AI Matching: AI analyses the data to match part-time workers with training programmes that align with their availability and career goals. For example, a part-time retail worker aiming for a full-time IT job might be recommended evening or online IT certification programmes that can be completed at their own pace.
3. Outcome: Part-time workers receive personalised training recommendations that accommodate their work schedules while helping them build skills for career advancement. This flexibility increases access to education for workers who otherwise might not have the time to pursue formal training, helping them transition into higher-paying or full-time roles.

Use Case 14: Boosting Rural Workforce Development

In this scenario, the D-ILA model is used by a regional government to boost workforce development in rural areas by identifying and matching workers with training that addresses regional industry needs.

Context: A regional government wants to develop the skills of its rural workforce to meet the needs of local industries, such as agriculture, tourism, or renewable energy. Many rural workers face barriers to accessing training due to geographic isolation and limited educational resources, and the government seeks to use ILAs to provide financial support for training programmes that cater to rural populations.

How the D-ILA Model Works:

1. Data Collection: The model collects data on the skills, employment history, and aspirations of rural workers, alongside the regional industries that are key to the local economy. It also integrates data on available training programmes that are accessible to rural populations, including online courses or programmes offered at regional community centres.

Funded by
the European Union

2. AI Matching: AI analyses the data to match rural workers with training programmes that align with local industry needs. For example, a farmer looking to diversify their income might be recommended courses in agri-business, sustainable farming, or rural tourism, while other workers might be directed toward training in renewable energy technologies.
3. Outcome: Rural workers are provided with personalised training pathways that align with local industry needs and their individual circumstances. This approach helps develop a skilled workforce that can drive regional economic development and support industries critical to rural economies.

Use Case 15: Enhancing Diversity and Inclusion in Training Programmes

In this scenario, a government or corporation uses the D-ILA model to promote diversity and inclusion by ensuring that underrepresented groups have access to personalised training programmes that address their unique challenges and career aspirations.

Context: A government or large organisation wants to ensure that underrepresented groups, such as women, minorities, and people with disabilities, have equal access to training and career advancement opportunities. The goal is to provide personalised learning pathways that account for the specific barriers these groups face, such as limited access to training, discrimination, or workplace inequalities.

How the D-ILA Model Works:

1. Data Collection: The model collects data on individuals from underrepresented groups, including their career goals, educational backgrounds, and the specific barriers they face. It also integrates data on training programmes designed to support diversity and inclusion, such as leadership programmes for women or accessibility-focused courses for people with disabilities.
2. AI Matching: AI analyses this data to match individuals with training programmes that align with their needs and help them overcome barriers to career advancement. For instance, a woman in a male-dominated industry might be recommended leadership training and mentorship programmes, while an individual with a disability might be directed toward courses that focus on accessible technologies or remote working skills.
3. Outcome: Underrepresented groups are provided with personalised training recommendations that promote inclusion and career development. This ensures that diversity is prioritised within workforce development efforts, helping to create a more equitable and inclusive labour market.

VIII. Annex

Please, find the Annexes on the project's website: www.kifu.gov.hu/d-ila

VIII.1 AI Run Reports

VIII.2 Commensurability matrix

Funded by
the European Union