

ELDS workshop, Bratislava, 7.11.2024

Horizon Europe project DisAI:

Improving scientific excellence and creativity in artificial intelligence and language technologies to fight disinformation

Marián Šimko, KInIT



Funded by
the European Union

Kempelen Institute of Intelligent Technologies is a non-profit research institute focusing on AI



- research and transfer of cutting edge AI to industry
- established in 2020 in Bratislava, Slovakia
- more than 50 researchers and research engineers (including PhD students) and growing



1st

in Slovakia

27

countries

35+

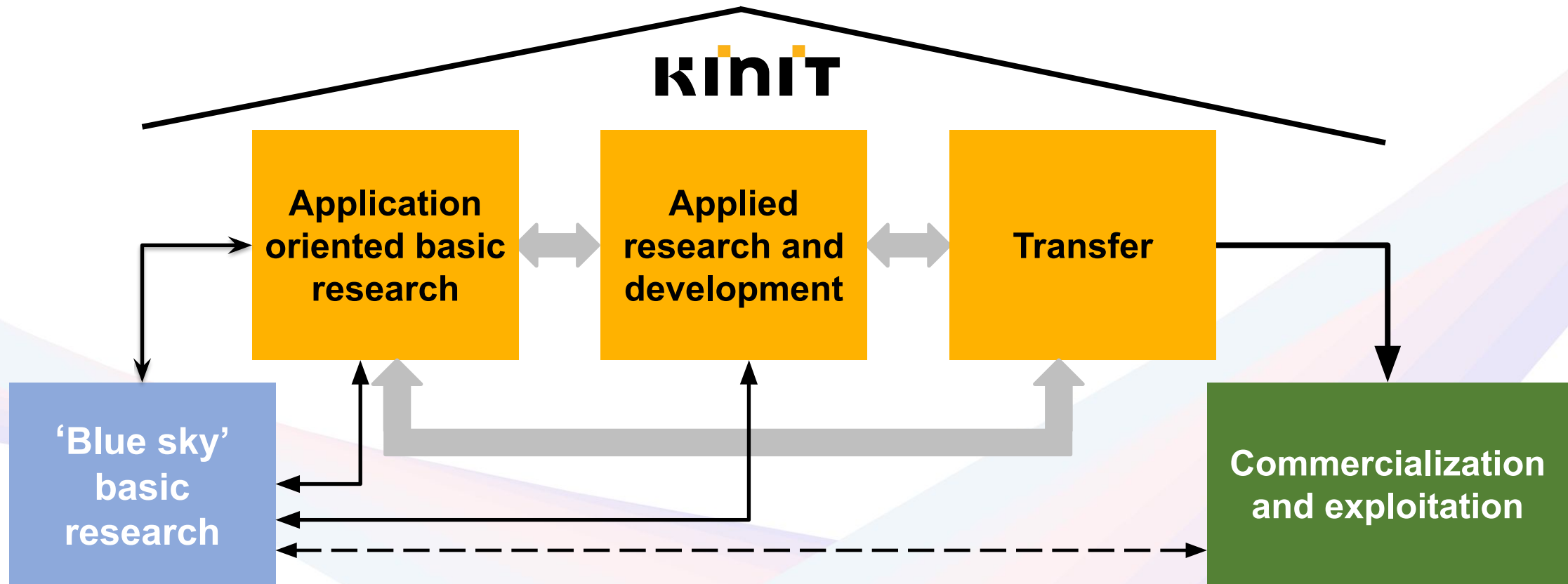
industry partners

The vision of KInIT is to accelerate the economic transformation of Central Europe



Bring together and nurture experts in artificial intelligence with connections to other disciplines

KInIT conducts basic research closely linked to applied research and knowledge transfer



Inspiration from abroad – we are not inventing the wheel

Three pillars of the non-profit science system in Germany

Universities
and
Colleges



Major
Research
Institutes



MAX-PLANCK-GESELLSCHAFT



SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN

DisAI objective is

to enhance the scientific excellence of KInIT and the consortium partners in

- trustworthy AI,
- multimodal natural language processing and
- multilingual language technologies

to combat disinformation.

CSA type of project (Coordination and Support Action) within the **Twinning** scheme

Specialty: **Research component** included

Project partners



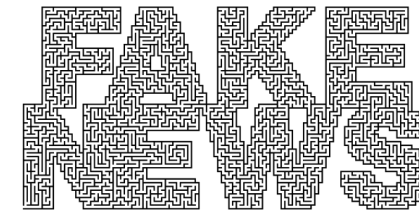
UNIVERSITY OF
COPENHAGEN



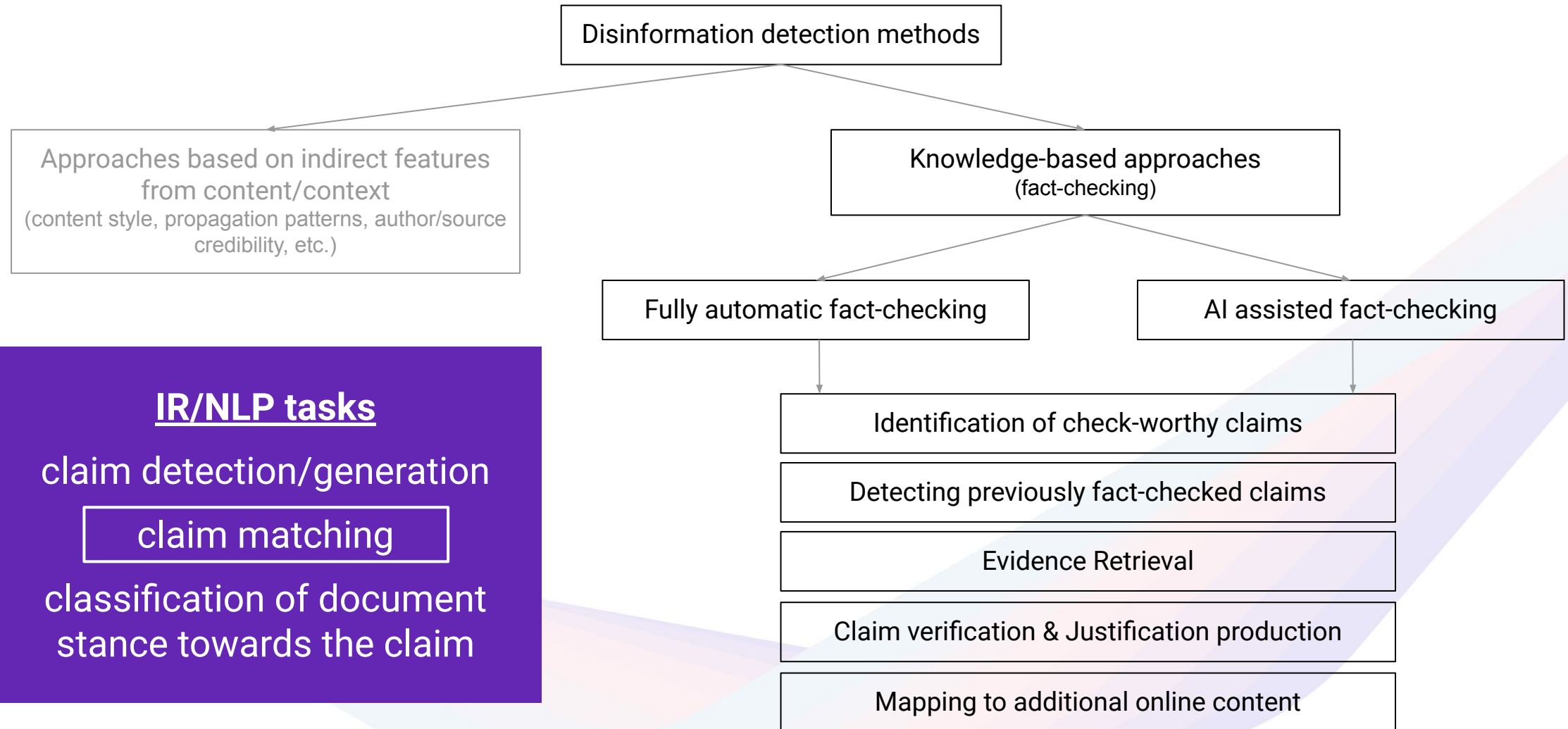
Disinformation and information disorders



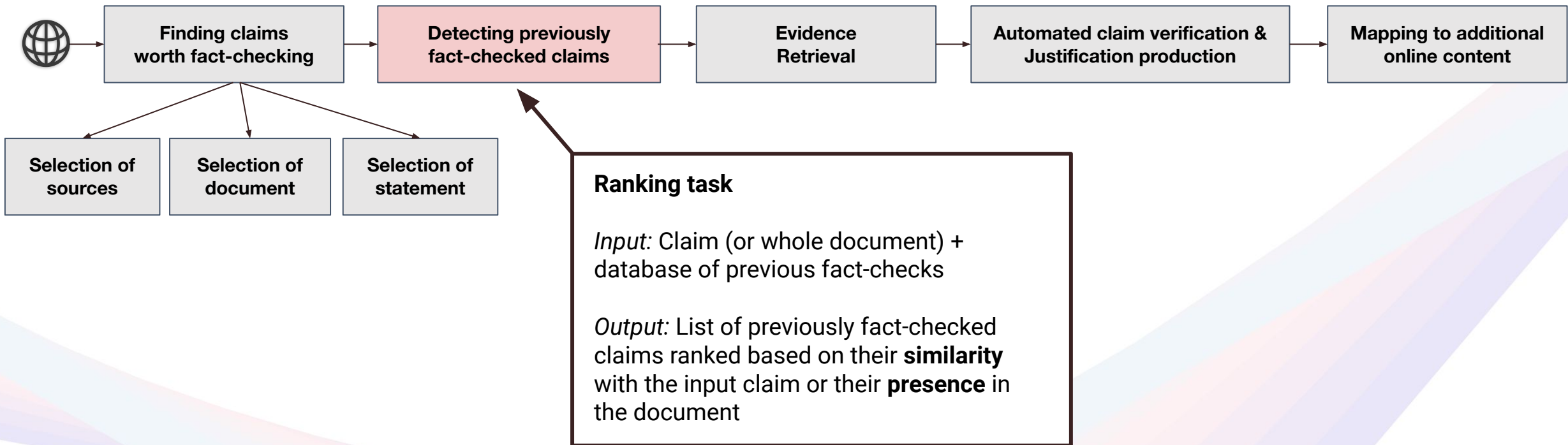
- a great risk for democracies and societies worldwide
- rising challenge in the era of social media
- recognized on the EU level
- problem far from resolved
- particularly important in Slovakia as a post-communist country
- artificial intelligence has potential to contribute to the fight against disinformation
 - verifying, identifying, tackling disinformation
- as well as it can be misused for disinformation spreading and generation



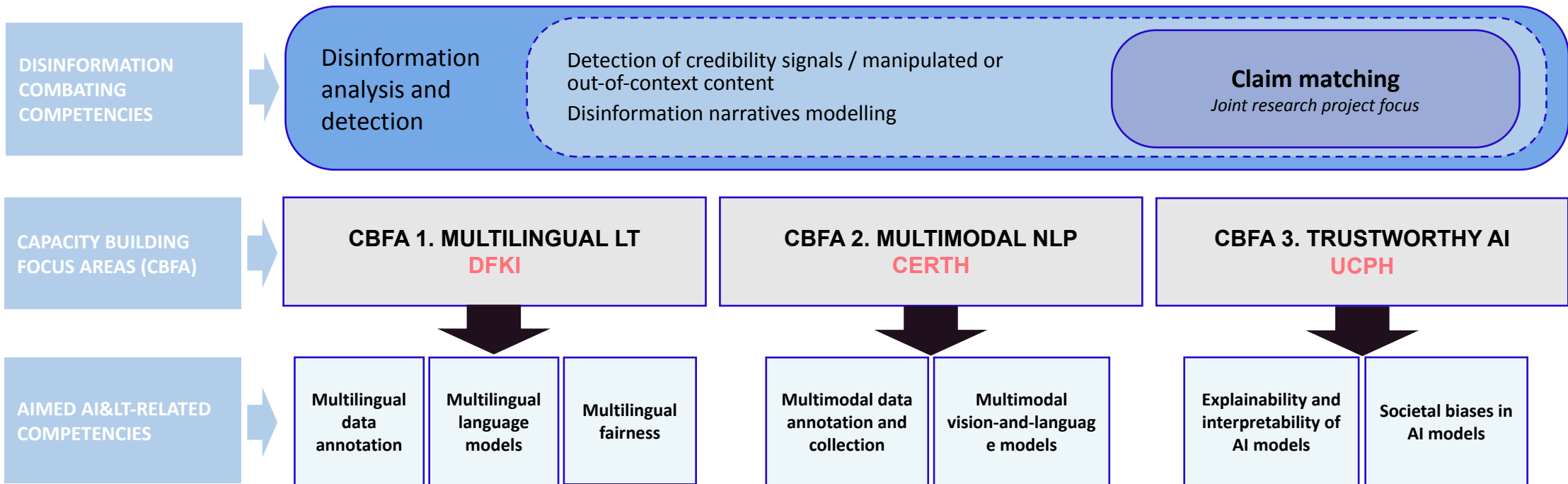
Approaches to disinformation detection



IR/NLP tasks
claim detection/generation
claim matching
classification of document
stance towards the claim



DisAI: Combating Disinformation with Artificial Intelligence and Language Technologies



Multilingual Language Technologies



- How can claim matching performance be improved by creating a multilingual machine learning model, i.e., a model using training data from multiple languages at the same time? Would there be a positive transfer of knowledge between languages?
- To what extent is it possible to perform cross-lingual claim matching, where a claim is in one language and the text we analyse in another? What are the limitations of typological similarity between languages in this case?
- How to perform fair evaluation of multilingual language technologies. How to make sure that the evaluation is fair w.r.t. low-resource languages and takes into consideration their needs.

Multimodal NLP



- Can we effectively analyse the visual component of the articles to improve the claim matching models? I.e., are we able to include the images from the articles in the model to increase its performance?
- Is it viable to use multimodal models to transfer knowledge between languages to improve the performance for low-resource languages?

Trustworthy AI



- How to explain the predictions of claim matching models understandably and objectively correctly to the users? Claim matching is a challenging task, and the interplay between two texts must be visualised.
- How can we use explanations to increase robustness of neural models?
- How can we detect and mitigate unwanted biases in claim matching models?

Rise of GenAI and LLMs changes the landscape



- LLMs are state-of-the-art for vast number of NLP tasks
- capabilities of LLMs for low-resource languages are limited
- we connected Slovak NLP community (so far, academic part, but working on)
- we started an initiative that aims to improve the large language models and their usefulness for the Slovak language
- one of our goals: fine-tune several foundation models on Slovak data

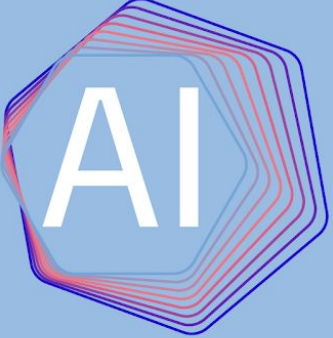
Most important Ingredient:
High-quality data in Slovak for different domains, tasks, ...
is still missing!

Conclusions



- Disinformation and information disorders are great risk for society
- For combating against them we can leverage AI and NLP
- Presence of suitable and efficient language technologies is crucial
- It is critical to develop and maintain high-quality language resources

Especially for low-/limited-resource languages like Slovak

Dis  .eu

kinit

Thank you for attention!



**Funded by
the European Union**

DisAI is funded from the European Union's Horizon Europe programme under Grant Agreement No 101079164.