



EUROPEAN LANGUAGE DATA SPACE



Common European Language Data Space /
Spás Sonraí Teanga Eorpach
The LDS Workshop in Ireland /
An Cheardlann LDS in Éirinn
Introduction / **Réamhrá**

Dr Brian Davis, Assistant Professor,
Ceann Cúnta Sármhaitheasa Teagaisc, School of Computing, DCU,
Imscrúdaitheoir Maoinithe agus Ceannaire Dúshláin Giniúint ADAPT
Principle Investigator - eStór

The Power of Artificial Intelligence (AI)

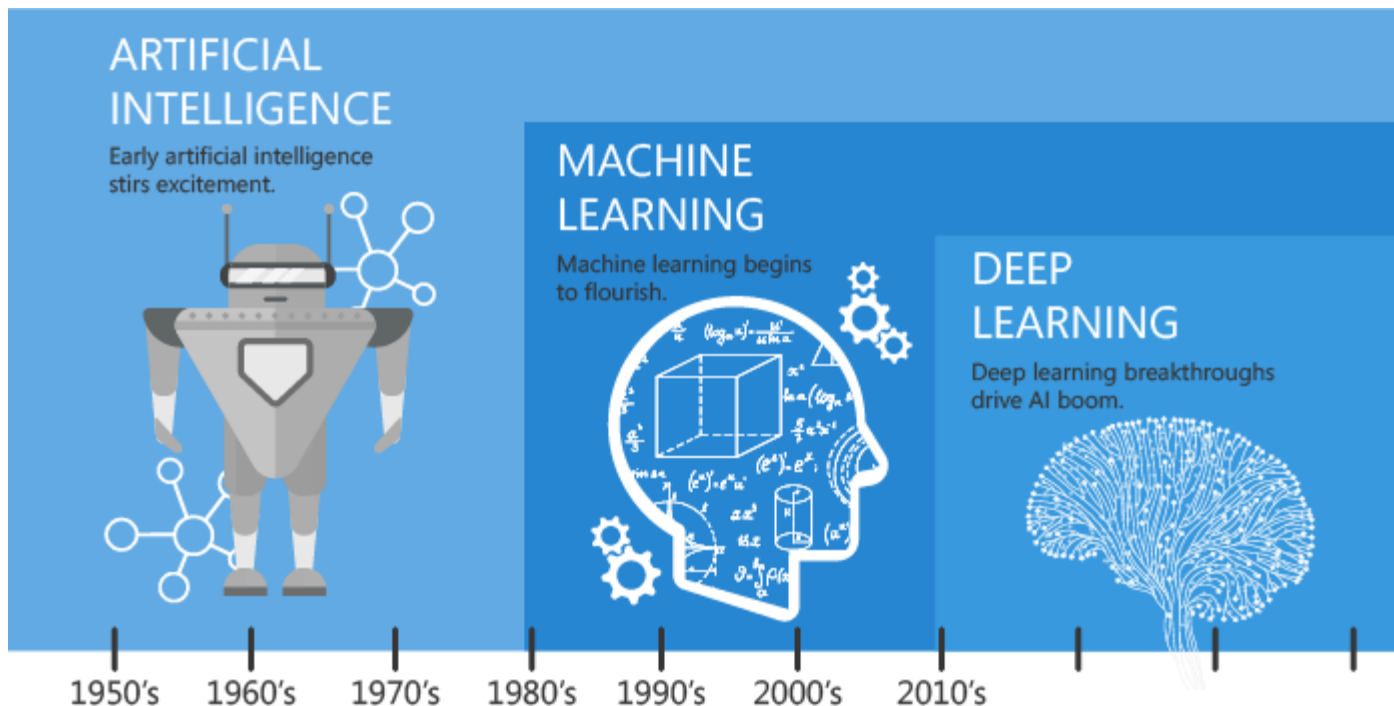
- AI represents one of the greatest opportunities for global societal and economic progress
- The National Artificial Intelligence Strategy: AI - Here for Good, was launched in July 2021 (*Notice no mention of AI for Irish*)
- Focus on AI in various new initiatives and projects across Europe, e.g.
 - AI4EU: First European AI On-Demand Platform and Ecosystem
 - Recent ICT-48-2020 Calls: Towards a vibrant European network of AI Excellence centres
 - Face2Face Virtual Agora: EU Artificial Intelligence Excellence Centres



Evolution of AI, Machine Learning, Deep Learning



Engaging Content
Engaging People



Definitions

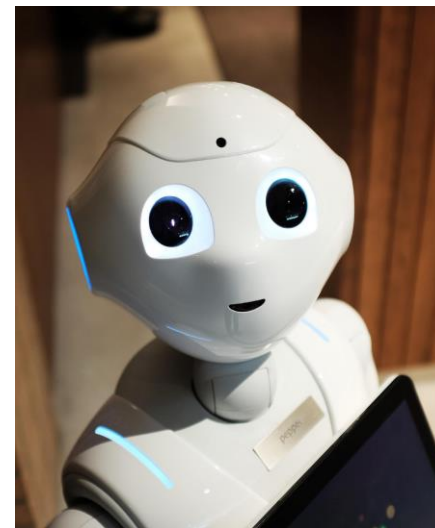
- **Language technology (LT)** — often also referred to as *human language technology* — comprises computational methods, computer programs and electronic devices that are specialized for analyzing, producing, modifying and translating text and speech



- **Machine translation** is a sub-field of NLP that investigates the use of software to translate text or speech from one language to another.

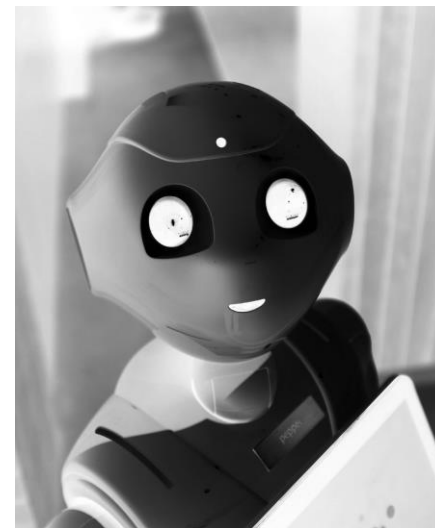
AI is required for many technologies in our lives

- Digital personal assistants
 - (e.g. Siri, Alexa)
- Smart vehicles
- Marketing
 - (e.g. Netflix, Amazon, Facebook)
- Health services
- Machine translation
- Dialogue systems / Chatbots
 - (e.g. in banking, management and customer support)
- Data analytics



AI is required for many technologies in our lives... BUT—

- Despite powerful advances which have seen widespread use in industry and public settings, ethical concerns with these technologies need to be addressed:
 - Bias in models discriminates against minority classes of society (including language, gender, race, sexuality, political or economic situation, etc.)
 - Data protection and IP concerns
 - While outputs can be convincing, they may be misleading or completely false
 - Shortsighted practises (e.g. ignoring environmental impact)
 - Lack of regulation regarding data collection, domain selection and model architecture
 - Bad actors



Why Human Language is so difficult to process

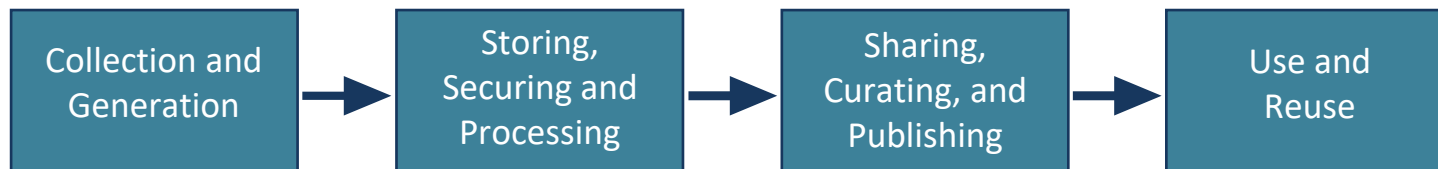
- 3 REASONS !
- AMBIGUITY!
- AMBIGUITY!!
- AMBIGUITY!!!



For AI to be effective, much more data is needed!



eSTÓR: Processing and Collecting Irish Data



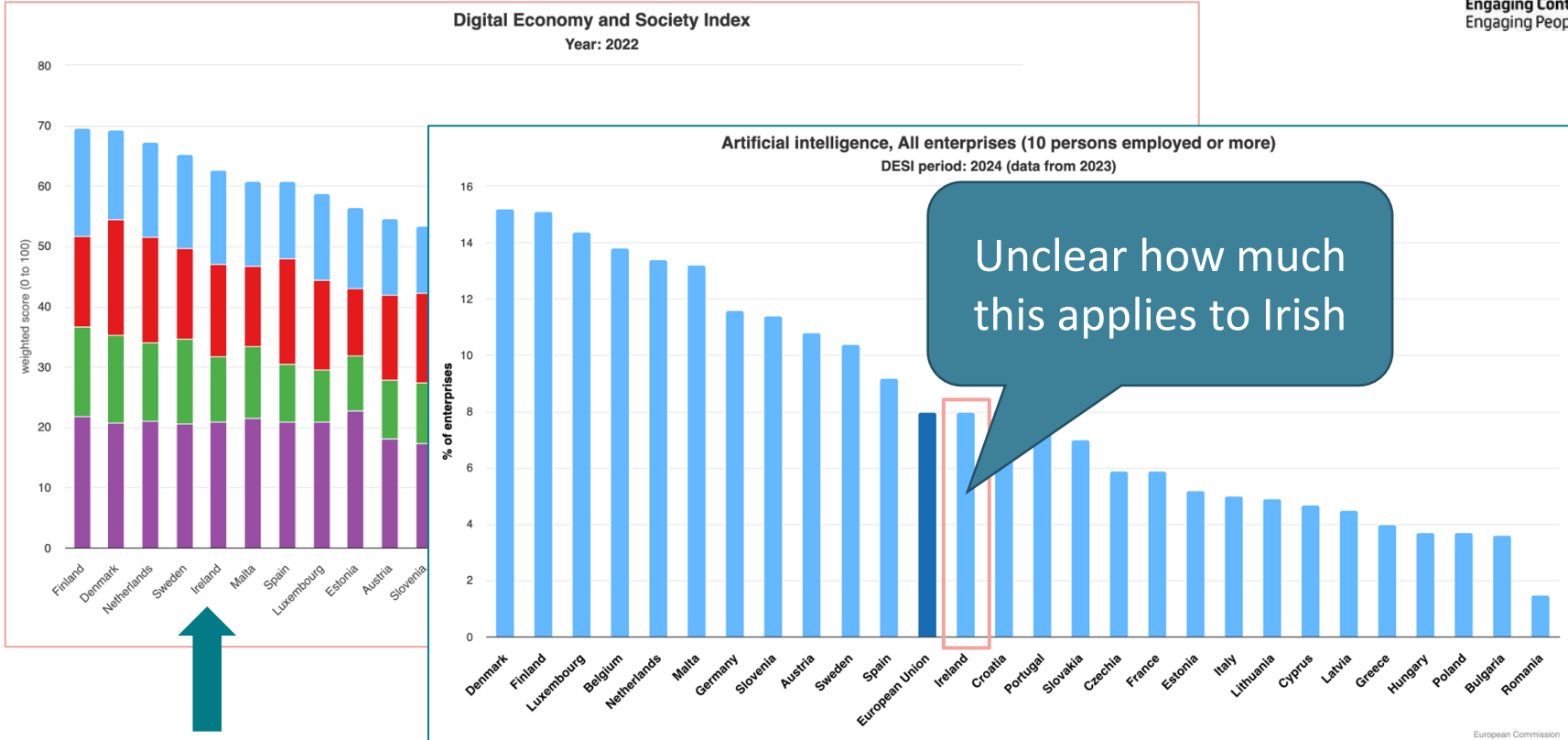
An Chartlann Náisiúnta
National Archives



Status quo: How Digital Are We?



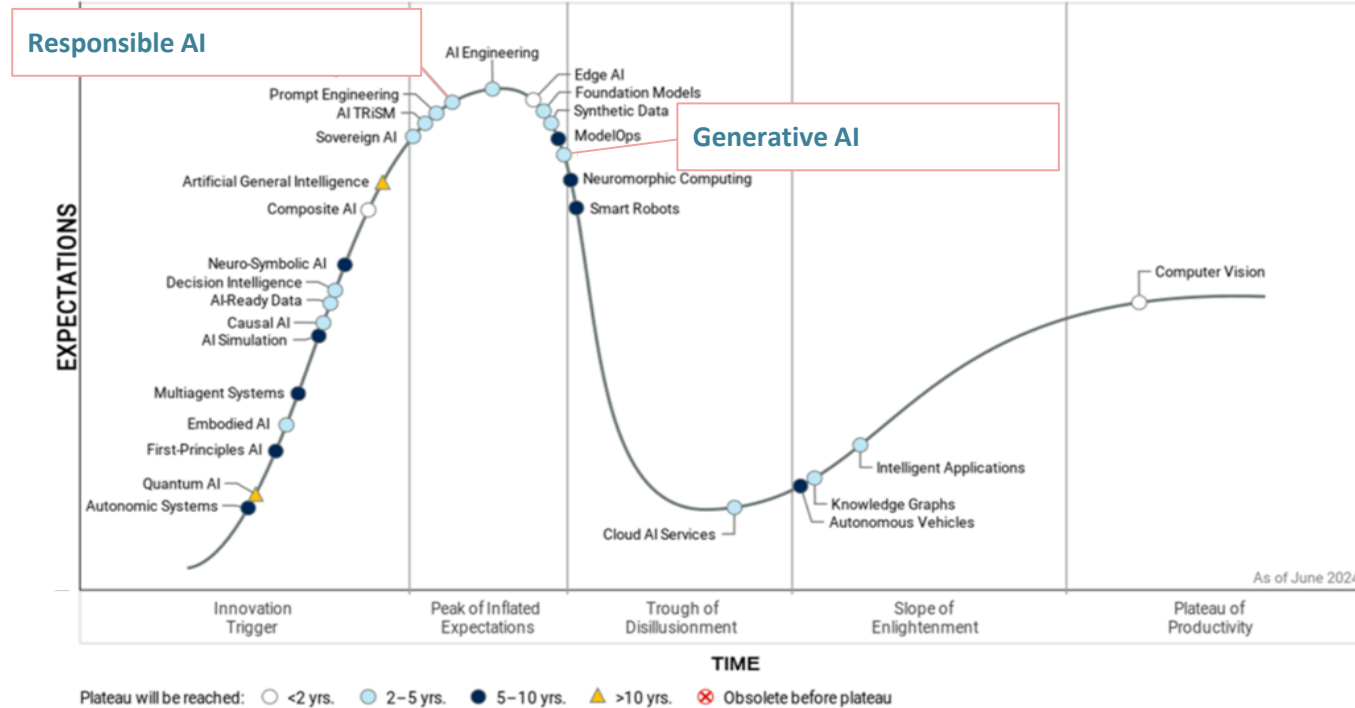
Engaging Content
Engaging People



Gartner's 2024 Hype Cycle for Digital Government Technology



Engaging Content
Engaging People



Source: Gartner (August 2024)



What Do We Need To Do?

- Develop effective data policies
- Encourage companies to pool and share their data
- Create data that is in the public interest
- Support the right to data portability
- **More domain specific data, especially for Irish**
 - e.g. legal Irish text, medical, public administration data, Gaelscoileanna

European Strategy for Data

- Data Act and European Data Governance Act
- Incentivise and facilitate companies to share their data through LDS
- Open Data Directive
- EU-funded projects through
 - Horizon Europe
 - Digital Europe programme
 - Connecting Europe Facility 2



Common European Language Data Space

Thank you and enjoy the workshop!



A Common European Language Data
Space – funded under contract LC-
01936389 with the European Union.

Dr. Brian Davis (ADAPT Centre, DCU)
brian.davis@adaptcentre.ie

10-10-2024 LDS Country Workshop, Dublin, Ireland
<https://language-data-space.ec.europa.eu>