

THE POTENTIAL OF LANGUAGE TECHNOLOGY AND AI – WHERE WE ARE, WHERE WE SHOULD BE HEADING

DR BRIAN DAVIS, ADAPT CENTRE,
DUBLIN CITY UNIVERSITY(DCU)



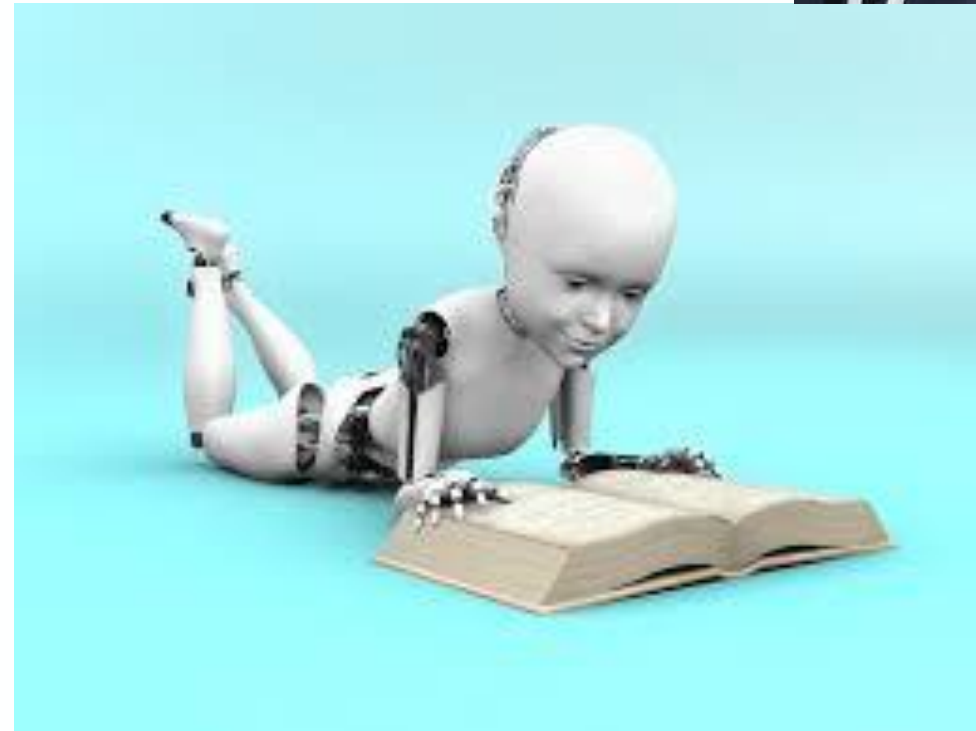
THE POTENTIAL OF ARTIFICIAL INTELLIGENCE (AI)

- AI represents one of the greatest opportunities for global societal and economic progress
- 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
- Already now there are very prominent examples of AI in our every day lives.....

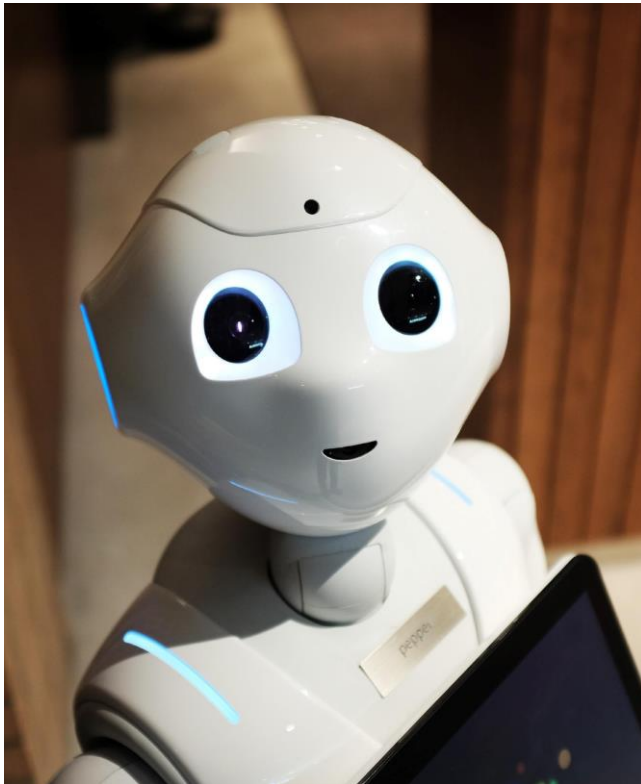
WHAT IS AI?

Merriam-Webster defines artificial intelligence this way:

- A branch of computer science dealing with the simulation of intelligent behaviour in computers.
- The capability of a machine to imitate intelligent human behaviour

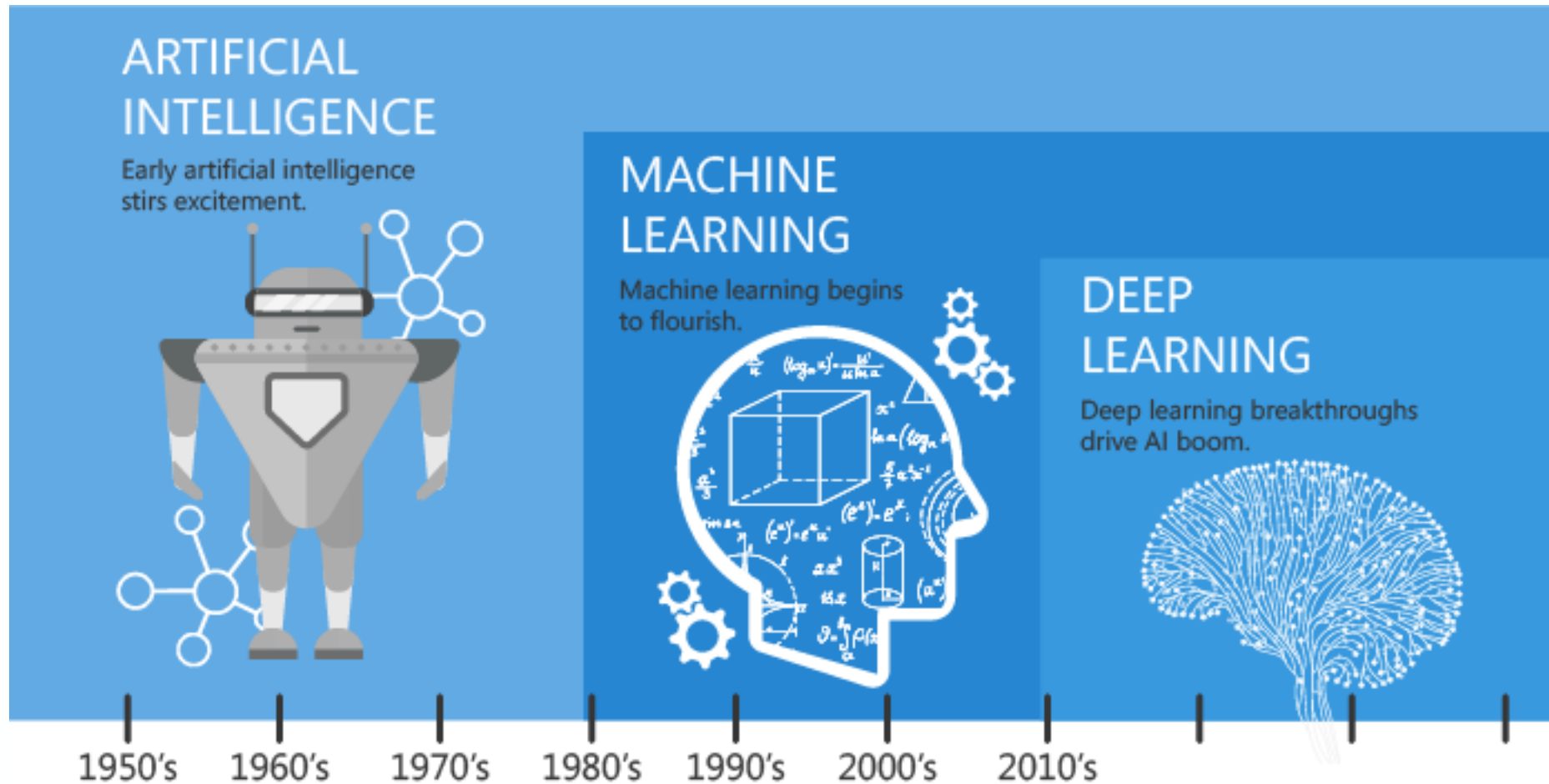


EXAMPLES OF AI IN OUR LIVES



- Digital personal assistants (Siri, Alexa, ...)
- Intelligent cars
- Marketing (see e.g. Netflix, Amazon etc.)
- Chatbots (e.g. in banking, customer support and management,)
- Finance (e.g. stock trading)
- Agriculture (optimal use of resources, population growth)
- Health care
- Space,

EVOLUTION OF AI, MACHINE LEARNING, DEEP LEARNING



Natural Language Processing(NLP) is the branch of Artificial Intelligence concerned with the study of the algorithms and data structures for processing human language.

Computational Linguistics (CL) is the branch of linguistics concerned with the scientific study of language from a computational perspective

NOTE: NLP and CL are in many ways flip sides of the same coin.



- **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. Not to be confused with computer-aided translation, machine-aided human translation or interactive translation.



WHY IS PROCESSING HUMAN LANGUAGE SO DIFFICULT?

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



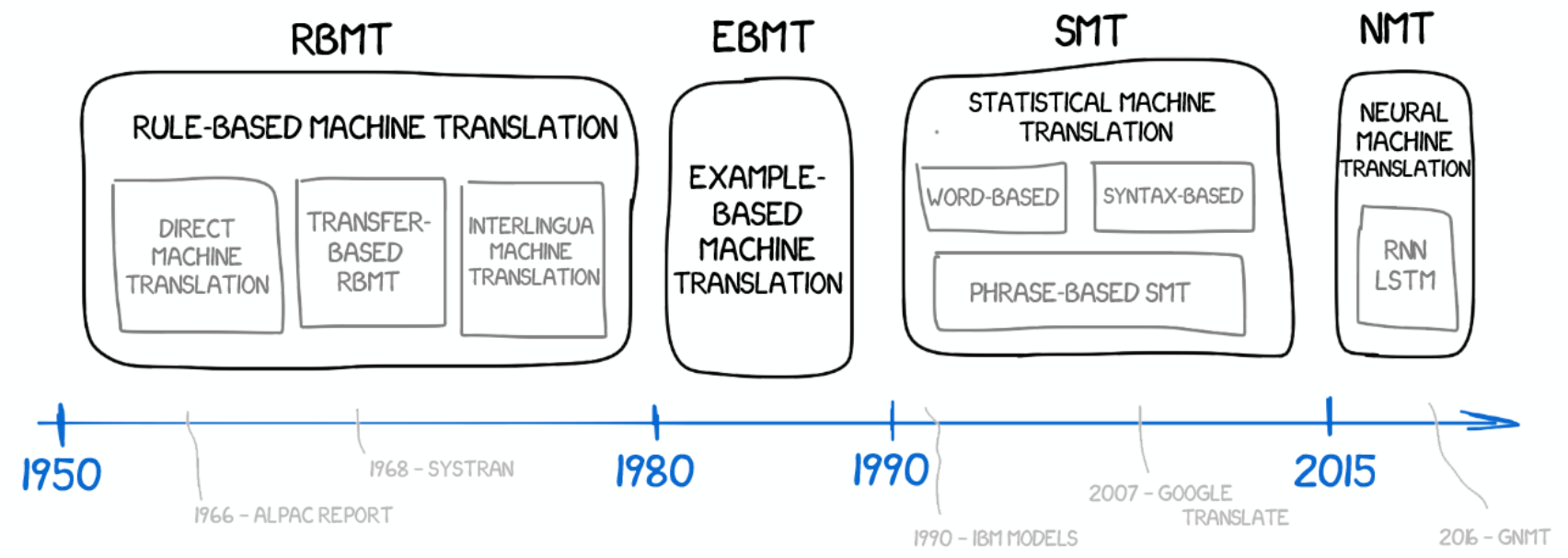
WHY IS PROCESSING HUMAN LANGUAGE DIFFICULT – EXAMPLES OF LINGUISTIC AMBIGUITY

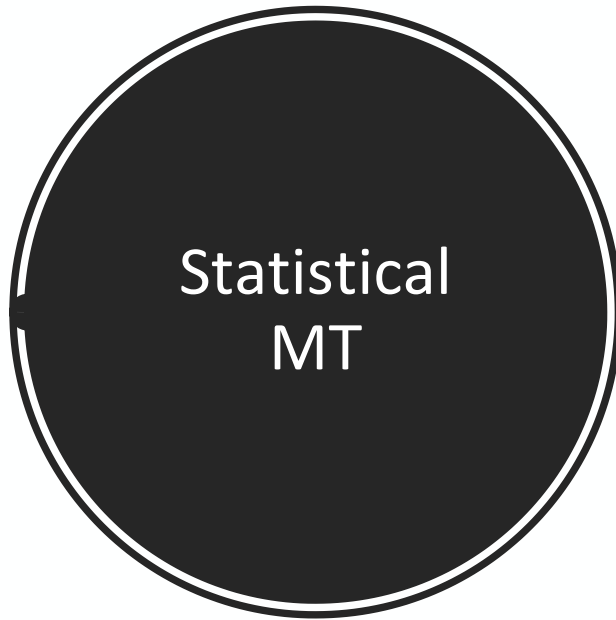
Some interesting examples...

- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Miners refuse to work after death
- Local High School Dropouts Cut in Half
- Red Tape Holds Up New Bridges

MACHINE TRANSLATION

A BRIEF HISTORY OF MACHINE TRANSLATION

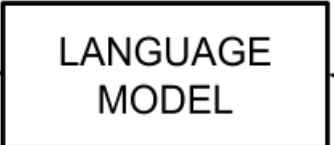




Ionchur Gaeilge



English output

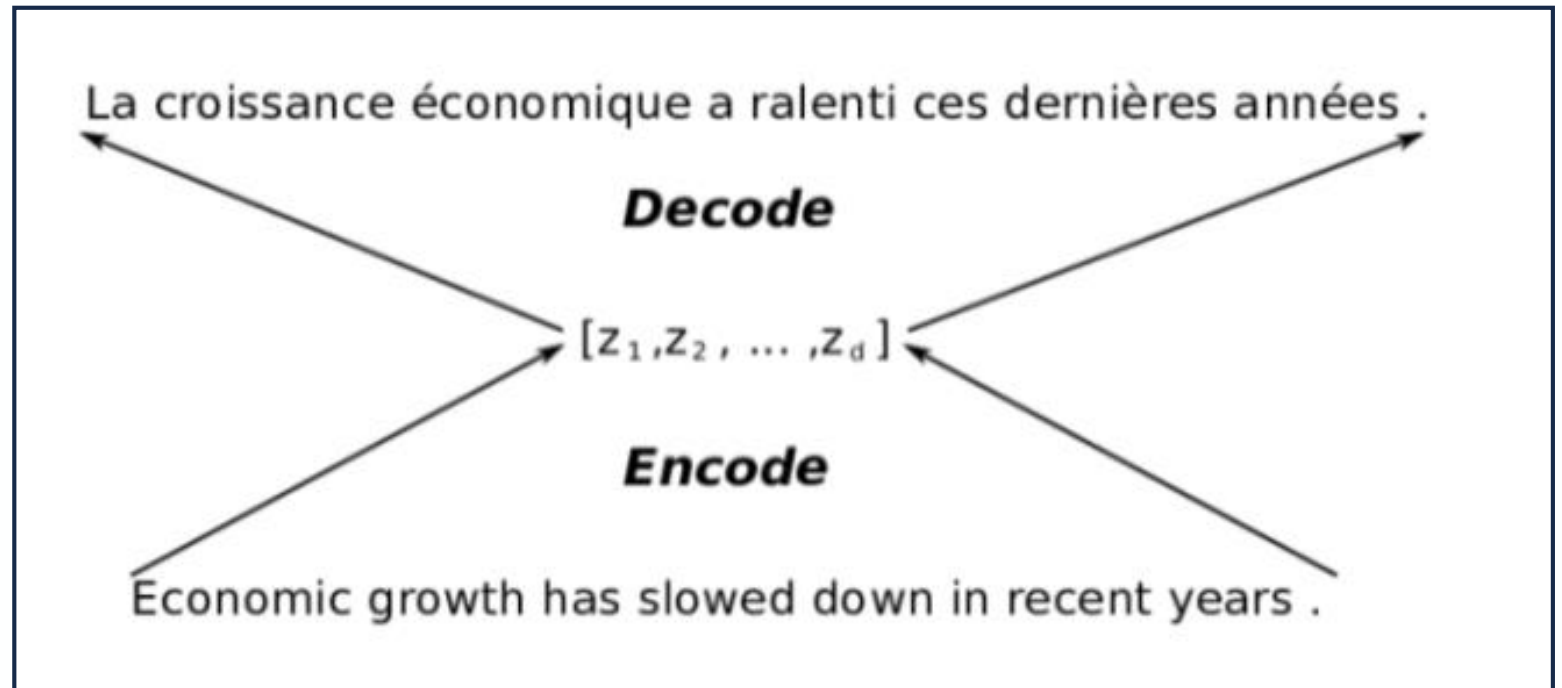
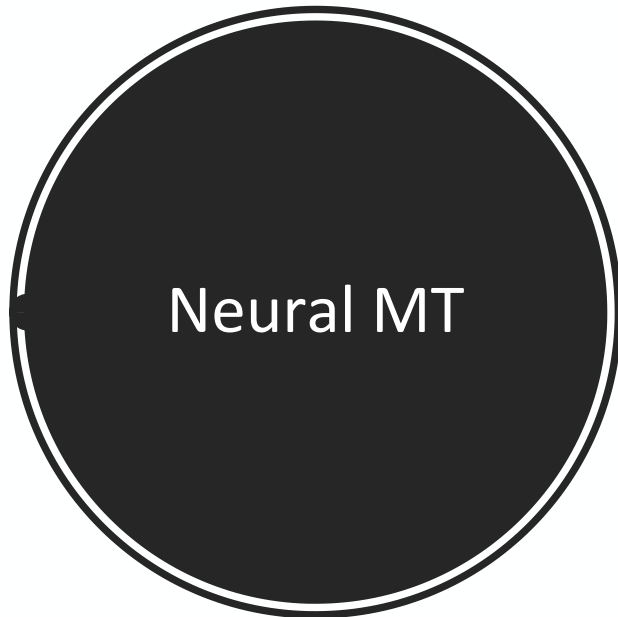


PARALLEL CORPUS

when you buy goods or services you enter into a contract with the supplier of goods and services .	nuair a cheannaíonn tú earraí nó seirbhís déanann tú conradh le soláthraí na n-earraí nó na seirbhísí sin .
when you engage a service supplier you create a contract between you , as a consumer , and the service supplier .	.nuair a théann tú i mbannaí soláthraí seirbhíse cruthaíonn tú conradh idir tú féin , mar thomhaltóir , agus soláthraí na seirbhíse .

MONOLINGUAL CORPUS

Is é rud is matamaitic ann ná staidéar ar phatrúin struchtúir, ar phatrúin athraithe, agus ar phatrúin spáis. Tagann an focal "matamaitic" ón nGréigis μάθημα (máthema), agus ciallaíonn sé "eolaíocht, eolas, nó foghlaim". Go neamh-fhoirmiúil d'fheadfá a rá gurb iad uimhreacha agus cló a hábhar staidéir. Dar leo siúd ar a nglaoítear na foirmiúlaigh, is taighde í de struchtúir teibí atá sainithe go haicsímiteach agus atá bunaithe ar loighic agus ar nodaireacht na matamaitice.



FOR MT TO WORK WE NEED LARGE AMOUNTS OF DATA!

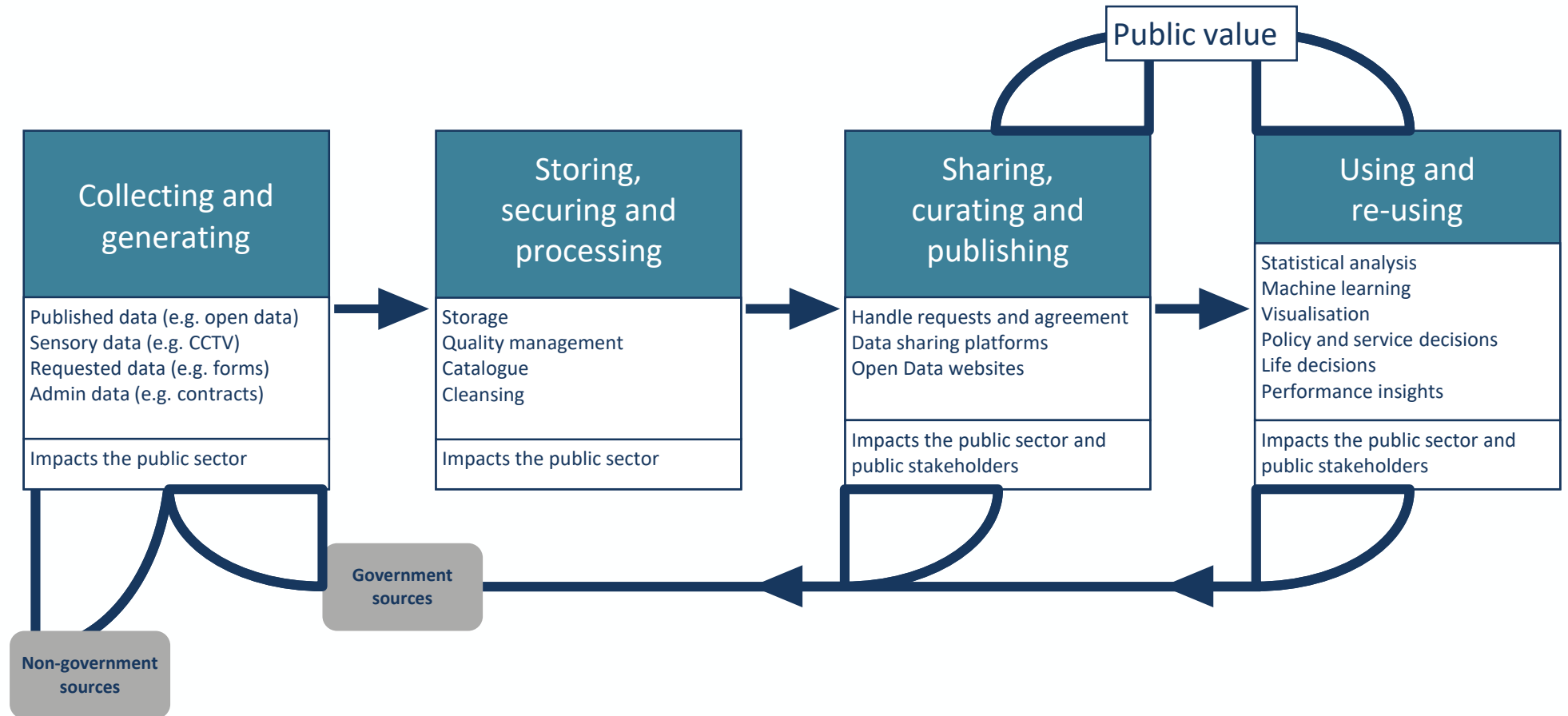


DATA DRIVEN LANGUAGE TECHNOLOGIES

AI applies to many other language technologies not just MTs, e.g.

- Acoustic Speech Recognition (ASR)
- Text to Speech Systems (TTS)
- Dialog Systems / Chat Bots
- Question-Answering Systems (QA)
- Named Entity Recognition (NER)
- Relation Extraction (RE)
- Text Summarisation
- Sentiment Analysis

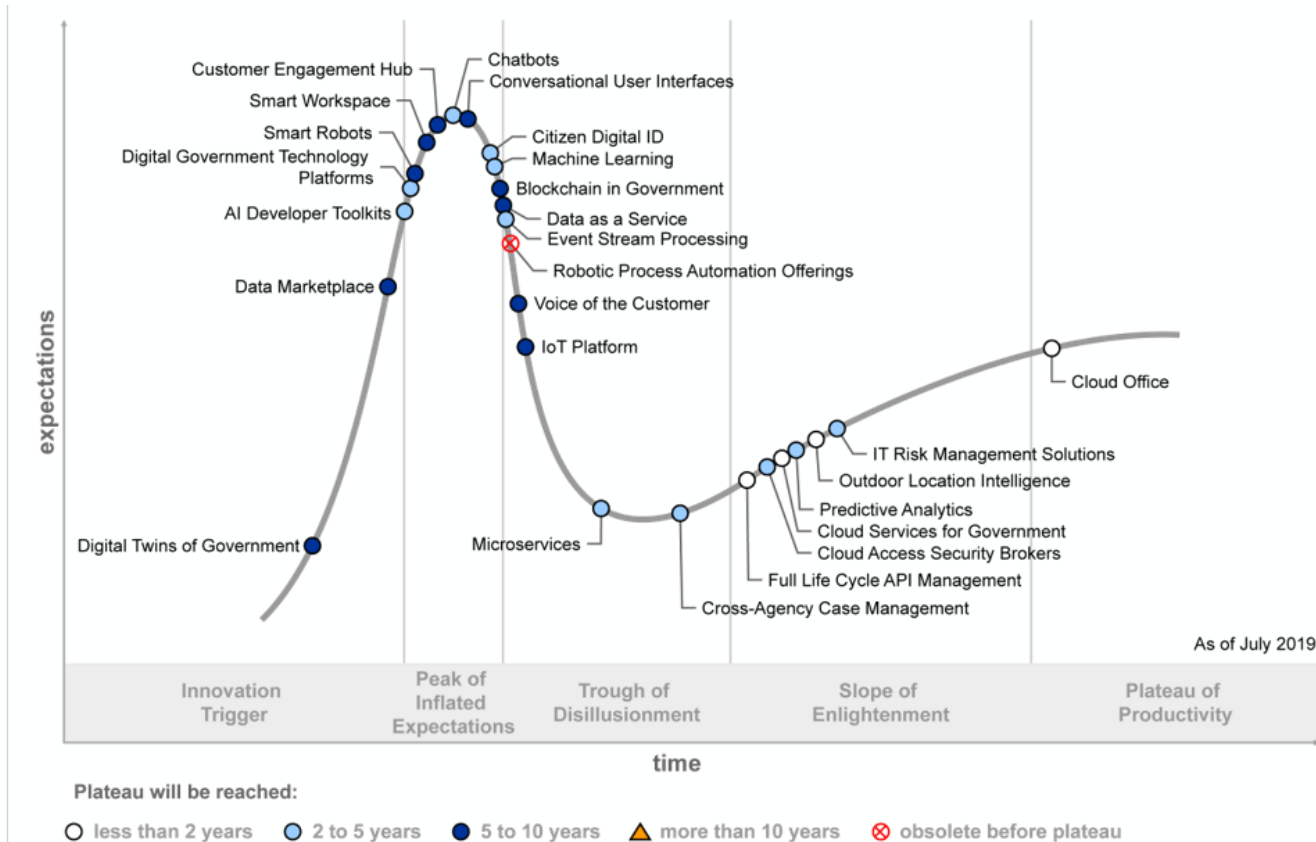
DATA AS FUEL FOR AI [



AI AND HLT – WHERE ARE WE HEADING TO?

Does NOT apply to
the Irish Language !

GARTNER'S 2019 HYPE CYCLE FOR DIGITAL GOVERNMENT TECHNOLOGY [5]



language technologies such as

- chatbot applications,
- voice solutions,
- conversational user interfaces,

are becoming increasingly important and will be mainstream in 5 to 10 years from now.

STATUS QUO: HOW DIGITAL ARE WE?

Digital Economy and Society Index (DESI)



IRISH LANGUAGE TECHNOLOGY – WHERE SHOULD WE BE HEADING TO?

AI: WHERE WE SHOULD BE HEADING?

- **Developing an effective data policy**
 - ☐ 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) i.e. legal Irish text, medical, public administration data.

THANK YOU FOR YOUR ATTENTION!

Website: www.lr-coordination.eu

Twitter: @LR_Coordination

Facebook: www.facebook.com/EuropeanLanguageResourceCoordination

LinkedIn: www.linkedin.com/in/lrcoordination

Email: info@lr-coordination.eu

