

HE Briefing Note Artificial Intelligence

March 2023



Introduction

Artificial intelligence (AI) is a branch of computer science that deals with the creation of intelligent agents, which are systems that can reason, learn, and act autonomously. AI research has been highly successful in developing effective techniques for solving a wide range of problems, including natural language processing, image recognition, and game playing.

Al is rapidly transforming the world around us, and education providers are no exception. Al is being used to improve teaching and learning, research, and administration. It is also being used to create new products and services that can benefit society.

ChatGPT is one AI tool that has dominated the news recently, though Google have just launched their competitor (Bard). These tools use GPT-3 and LaMDA as their language models respectively. Both models are capable of generating text that is indistinguishable from human-written text, and both were used in the drafting of this paper. These two models (and ones similar) will drive AI developments for the next few years.

There are some key differences to note between the two. We highlight the differences here to show that both models are very complex and well developed, and to illustrate how nuanced development of AI tools is becoming:

- 1. Purpose: GPT-3 is designed to generate text based on the input it receives, while LaMDA is specifically designed for dialogue applications, meaning it is optimised for generating responses to specific prompts in a conversational format.
- 2. Training data: GPT-3 was trained on a diverse range of text data, including web pages, books, and articles, while LaMDA was trained on conversation data specifically curated to mimic the types of conversations that people might have with Google Assistant.

- 3. Model size: GPT-3 is currently the largest publicly available language model, with over 175 billion parameters, while LaMDA is still in development and its size has not been publicly disclosed (though we expect it's bigger).
- 4. Architecture: GPT-3 uses a transformer architecture with multiple layers of processing, while LaMDA uses a more complex architecture that involves multiple types of neural networks, including transformers and recurrent neural networks (RNNs). In layman's terms, this means LaMDA is a more sophisticated tool (using more calculations in the process).
- 5. Applications: While both models can be used for a variety of applications, GPT-3 is more commonly used for tasks such as language translation, text completion, and text classification, while LaMDA is designed specifically for conversational AI applications, such as chatbots and voice assistants.

Benefits of AI for HEIs

There are many benefits of AI language models for providers, including:

- 1. Personalised learning: Al language models can be used to generate personalised learning materials for students, based on their individual needs and interests. This can help students to learn more effectively and efficiently.
- 2. Automated tasks: Al language models can be used to automate tasks such as grading papers, providing customer service, and providing feedback to students. This can free up staff time for other tasks.
- 3. Research: Al language models can be used to conduct research, for example by analysing large amounts of data or by generating new hypotheses. This can help providers to advance knowledge and understanding.
- 4. Innovation: AI language models can be used to develop new products and services, for example by generating new ideas or by

creating new ways to communicate. This can help providers to stay ahead of the curve and to remain competitive.

5. Enhanced communication: AI language models can be used to enhance communication between providers and their stakeholders, for example by providing real-time updates or by generating personalised messages. This can help to build trust and relationships.

Risks of AI for HEIs

There are also some risks associated with the use of AI language models for providers, including:

- 1. Data privacy and security: Al language models are trained on massive amounts of data, which could pose privacy and security risks if that data is accessed by unauthorised individuals.
- 2. Bias: Al language models are trained on data that is biased, which could lead to the generation of biased text (for example, GPT-3 is based on textual data taken from the internet, where progressive viewpoints are often more vocally represented, so these viewpoints may reappear within the responses produced).
- 3. Unethical use: Al language models could be used for unethical purposes, such as generating fake news or propaganda, or more relevant to the sector, academic misconduct.
- 4. Misuse: AI language models could be misused, for example by generating hate speech or by creating deepfakes.

It is important for providers to be aware of these risks and to take steps to mitigate them. For example, providers can:

- Use AI language models in a secure environment.
- Train AI language models on data that is not biased.
- Monitor the use of AI language models to identify any potential risks.

- Develop ethical guidelines for the use of AI language models.
- Educate staff and students about the risks of AI language models.

Marking the Turing Test

One of the common current concerns in the sector is that academics marking essays, will now need to 'mark' the Turing test, and assess whether AI language models have been used to generate a response. The Turing test is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. In the test, a human evaluator engages in a natural language conversation with two other parties: a human and a machine designed to generate human-like responses. During a text-based conversation, if the evaluator in the diagram below cannot reliably tell the machine from the human (Turing originally suggested that the machine would convince a human 30% of the time after five minutes of conversation), the machine is said to have passed the test.



The Turing test is frequently used to measure the progress of artificial intelligence. However, it has been criticised for being too anthropocentric, and for being too easy to game. Nevertheless, it remains a popular measure of artificial intelligence.

There are some key pointers that academics can look out for in responses to try and identify the use of AI models:

1. Consistency: AI-generated responses are often highly consistent in terms of style and tone, struggling to mimic the natural syntax and

grammatical flow that may be produced by individuals. They may also repeat certain phrases or use specific vocabulary more frequently than a human would.

- 2. Unusual responses: AI models may generate responses that are technically correct but sound unusual or unnatural to a human ear. Humans may also detect errors or inconsistencies in the AI-generated responses.
- 3. Speed and volume: AI models can generate responses very quickly and in large volumes. If a response is generated almost instantly, or if there are a large number of responses generated in a short period of time, it may suggest that they were generated by AI.
- 4. Knowledge limitations: AI models may struggle with complex or abstract topics that require nuanced understanding, creativity, or critical thinking. If an answer seems to lack depth or originality, or if it relies heavily on surface-level information, it may suggest that it was generated by AI.

Alternatively, providers could adapt their exam strategies, using one or more of the following:

- Diagrammatic or visual questions that can't be used as an input in Al language models.
- Live examinations and presentations, enhanced by follow up questions which limit the use of AI.
- Ensuring examination focuses on application of learnt knowledge / references to course study rather than regurgitation of information.

Some in the sector advocate for further 'gamification'¹ of the higher education experience, but in some respects, this is a double-edged sword. The key aspects of gamification (competition / reward) are what drives students to use tools such as AI language models in the first instance.

Keeping a finger on the pulse

Here are some ways that higher education providers can get the most out of AI, while limiting the associated risks:

- 1. Use AI for good. AI can be used to dramatically improve the staff and student experience, and to make higher education more accessible. For example, AI can be used to personalise learning, to automate tasks, and to provide support services.
- 2. Be transparent about how AI is used. Higher education providers should be clear about how they are using AI and should make sure that students and faculty understand how AI works and how it is being used.
- 3. Protect student privacy. Al systems can collect a lot of data about students, so it is important to protect student privacy. Providers should have strong privacy policies in place and should make sure that students understand how their data will be used.
- 4. Monitor the use of AI. Providers should monitor the use of AI to make sure that it is being used in a responsible way. For example, AI should not be used to discriminate against students, or to make decisions that could harm students, and should not be used to elicit academic dishonesty.
- 5. Be prepared for the future. Al is a rapidly evolving field, so it is important for providers to be prepared for the future.

¹ Gamification is the strategic attempt to enhance systems, services, organisations, and activities by creating similar experiences to those experienced when playing games in order to motivate and engage users.

We can help

GPT-3 and LaMDA were used in the initial draft of this briefing paper, demonstrating a potential use case for AI language models in the sector. We can help provide insight across your institution where these may be of use, and how they would be best implemented. If you want to discuss any of this briefing paper in further detail, please get in touch.



Paddy Marshall Assurance Consultant e: pmarshall@uniac.co.uk <u>www.uniac.co.uk</u>

