Emerging trends and Applications in Artificial Intelligence

Assoc. Prof. Grant Dick, Head of School, School of Computing, University of Otago

A view to the past …

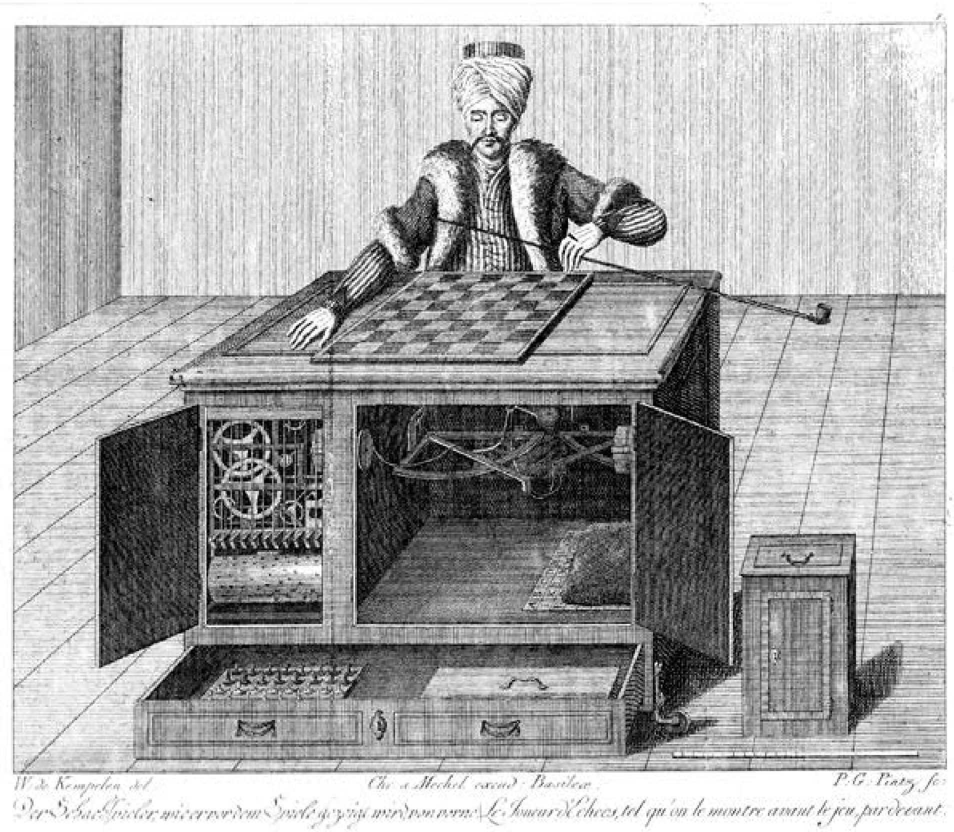
… c.1770 to be exact!

The “Mechanical Turk”
(c.f., Amazon Mechanical Turk)

Was promoted as an automaton capable of expert chess playing

Reality: elaborate ruse! Human expert hid inside the “machine” and controlled it.

(We’ll revisit this later!)



Artificial Intelligence (AI)

“AI is the science of making machines do things that would require intelligence if done by ~~men~~humans.”
- Marvin Minsky (“Father” of AI)

Essentially, “AI” is the process whereby machines/software/systems are developed to respond autonomously to given input in a productive manner.

Originally, AI was a philosophical endeavour, but has increasingly moved into an engineering sphere with a focus on **building systems** constructed upon **data-driven models developed through machine learning**.

AI vs. “Machine Learning”

Machine learning is the construction of predictive models through use of algorithms working on data

AI uses machine learning, among other things, to achieve its tasks of autonomous behaviour

Examples:

- “Find all the street signs in this image” is an example of machine learning
- “Develop a self-driving car that recognises and responds to street signs” would be an AI.

AI vs. “Data Science”

Important to distinguish the two.

- AI emphasises automation of the machine (autonomy without explanation)
- Data Science is about using statistical methods (incl. AI and ML) to extract meaning an understanding from data sources (essentially, intelligence gathering)

Examples:

- An AI might be developed to automatically triage/approve loan applications based on application details
- Data Science would be used to understand **WHY** people default on their approved loans

(both would use similar data sets!)

Modern “AI” is really an integration of all these things (systems, machine learning, and data science)

Types of AI

Operational (Task-Oriented) AI – applying intelligent methods to achieve a specific goal/task.

(note: **PLENTY OF EXAMPLES!**)

Artificial General Intelligence (AGI) – general task solving through machines that can learn/adapt

(note: **THIS DOESN'T EXIST YET!**)

Generative AI (should really be called *Imitative AI*) – generation of "sequences" (text, image, sound, video, …), often in response to a **prompt/stimulus**.:

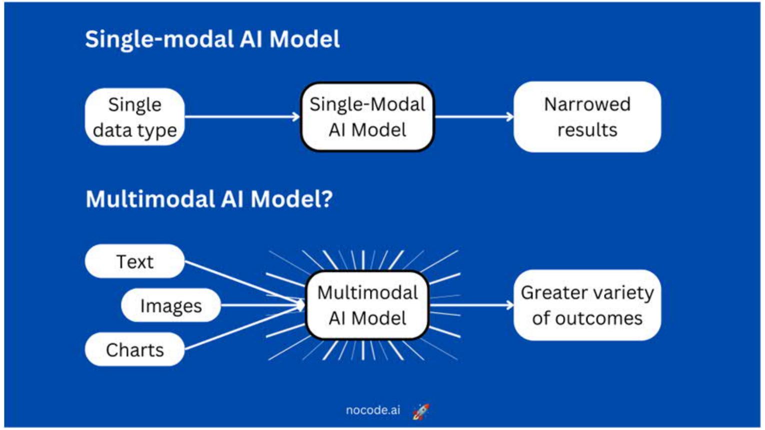
- Examples of this would include ChatGPT and DALL-E

Multimodal AI

AI that integrates multiple sources of information (e.g., text + video + tabular) as input.

Multimodal AI also typically responds with multiple types of output

Most current interest in AI is in **multimodal generative AI**



Large Language Models (LLM)

Essentially, LLMs are large statistical models that determine the probabilities for the next token in a sequence for a given context

(in the case of text: “the next word in sentence following a prompt”)

Typically, very large – **billions of parameters, trained on billions of tokens**. Require very large infrastructure for training.

Typical approach is to refine an existing LLM with local data rather than build from scratch (called **transfer learning**)

Note: ChatGPT (the AI) is not just an LLM, it **uses an LLM** (and other things) within its infrastructure **(more on this later)**

What is the fascination with LLMs (and related models?)

LLMs have very high capacity

Results in many useful applications

- Text summarisation
- Chatbot generation
- Assignment writing
- Object detection
- Source code generation
- …

Has led to many (erroneous, IMO) negative claims around the future of work and what employment looks like

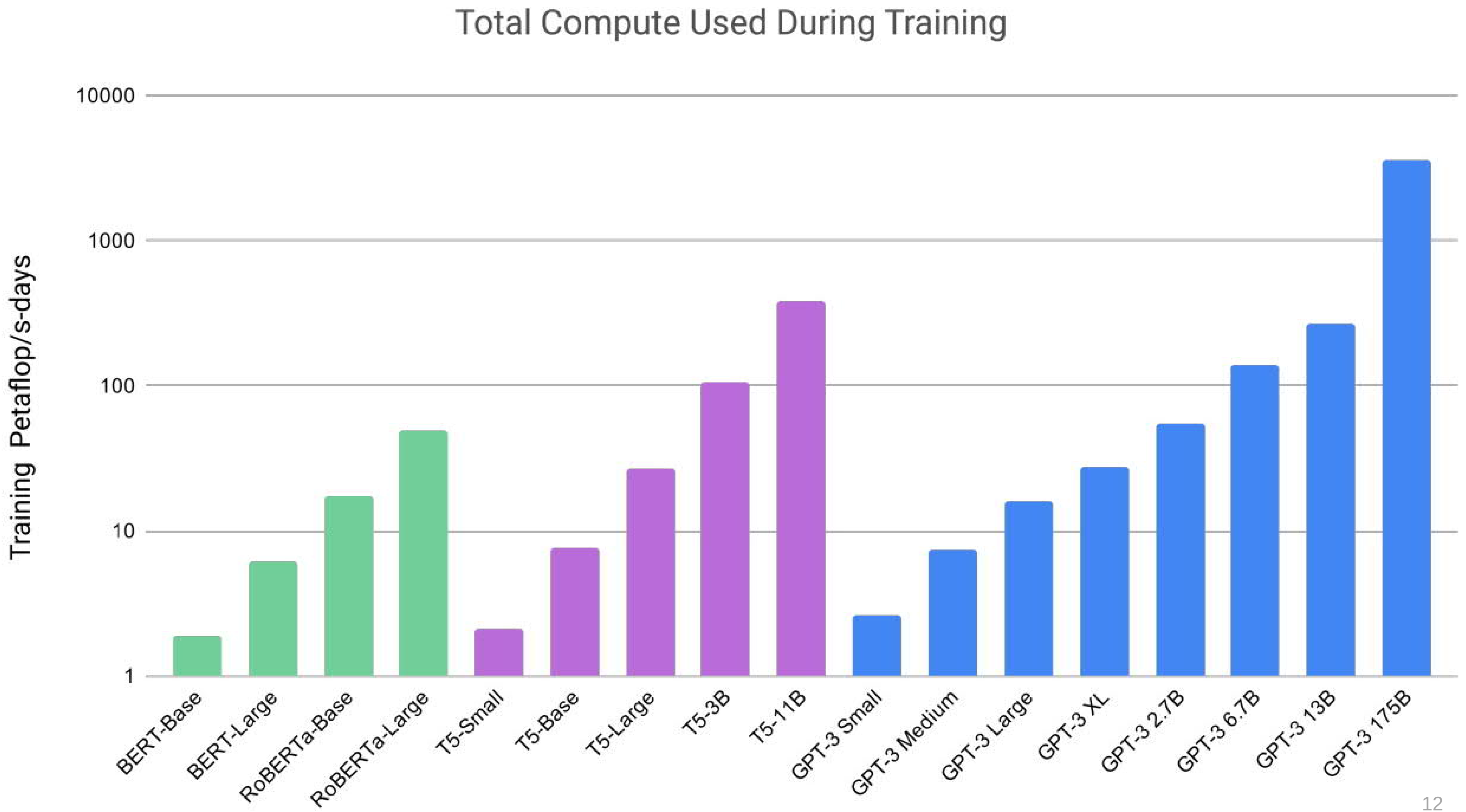


Image Source: Language Models are Few-Shot Learners (arXiv)

Foundation Models

LLMs are difficult and costly to train, but easy to adapt

Adaptation typically requires less computational resources than retraining from scratch

Current focus in (vendor-driven) AI is presentation of pre-trained foundation models

Foundation models typically accessed through vendor API

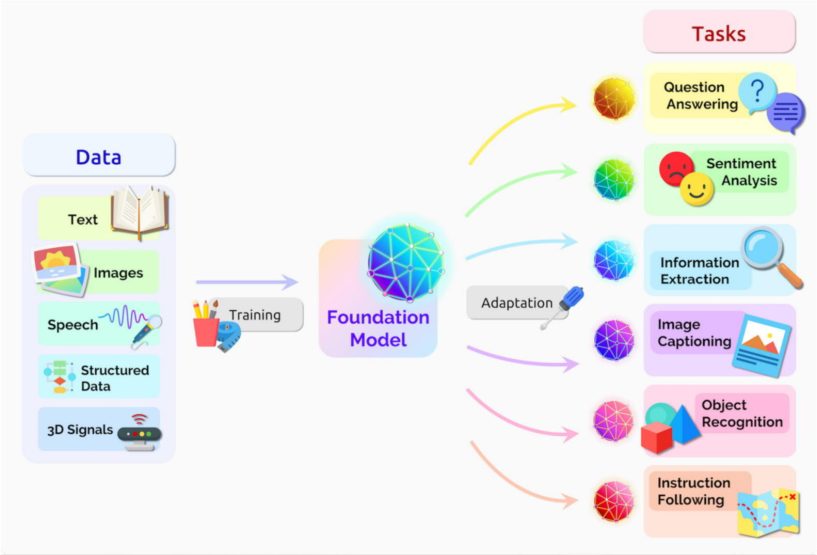
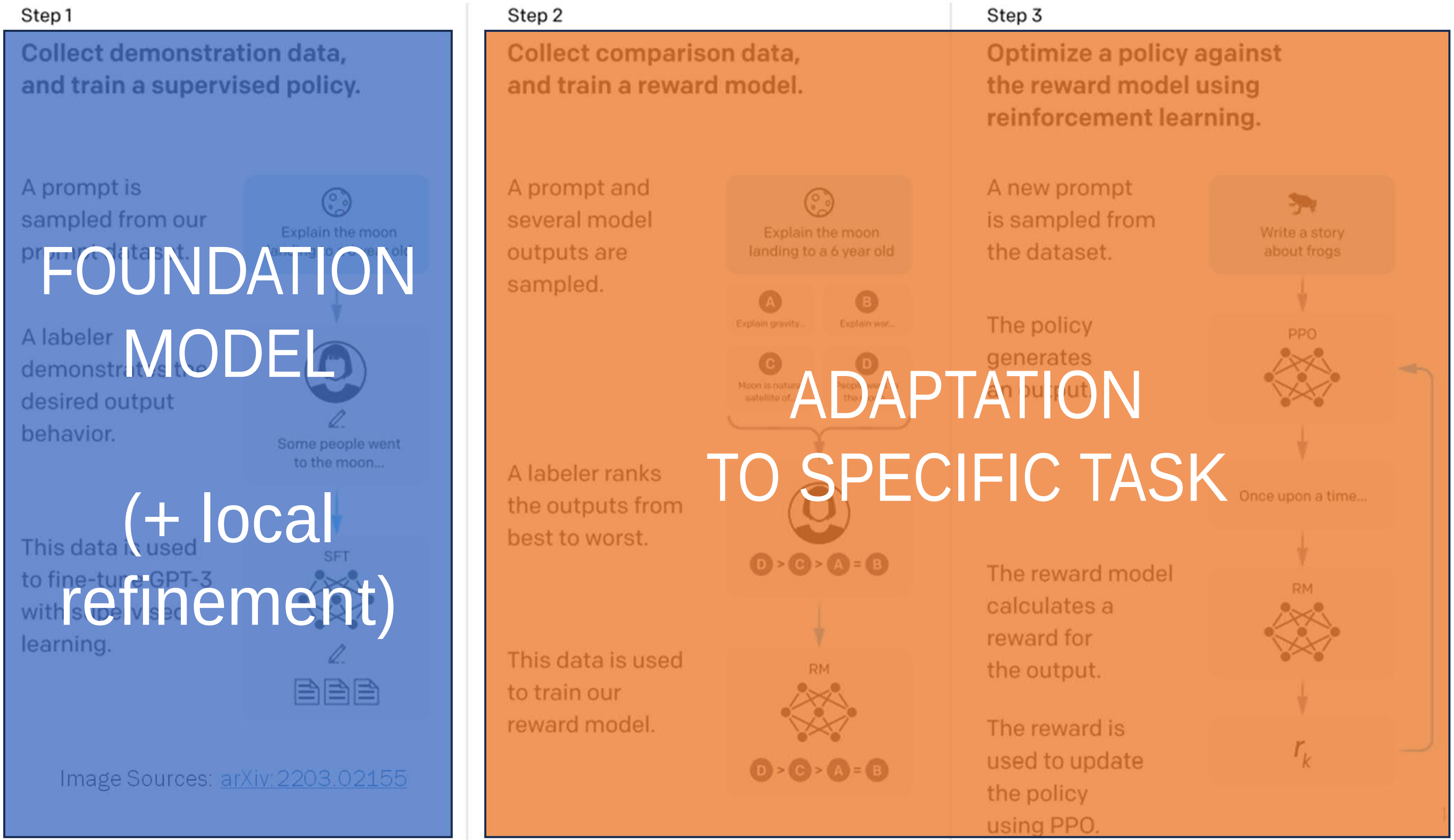


Image Source: On the Opportunities and Risks of Foundation Models (arXiv)



FOUNDATION MODEL

(+ local refinement)

ADAPTATION TO SPECIFIC TASK

Image Sources: arXiv:2203.02155

Examples

That are (maybe?) NZFSSRC relevant?

- Multimodal AI for risk assessment
- Practical-time methane source prediction from satellite imagery (surrogate vs. mechanistic models)
- Predict crop/stock/environment impact in response to natural disaster/event
- Translate risk models into interpretable and actionable knowledge for stakeholders
- …

Concerns with AI usage?

- Privacy and ethics
- "Intelligent" ≠ "Smart"

  often the response **lacks context**
- "Intelligent" ≠ "Aware"

  AI does not "know" what it is producing

  (hallucinations)
- Models quickly become out of date

  (anachronistic)
- Results are often not deterministic

  (same query gets a different response)

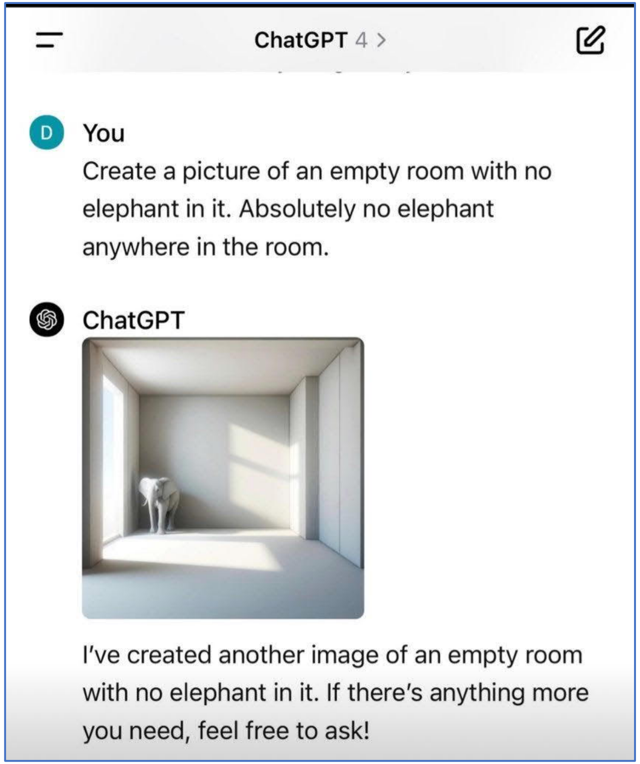


Image Source: X (@neerjkvaid)

Solutions?

Explainable AI (XAI)

Modern AI models large and "complex" (black boxes):

- Often focused on **WHAT** to predict
- Practical adoption limited through (valid) fear, caution and lack of transparency in actions

XAI attempts to augment model predictions with metadata about **WHY** the prediction was made

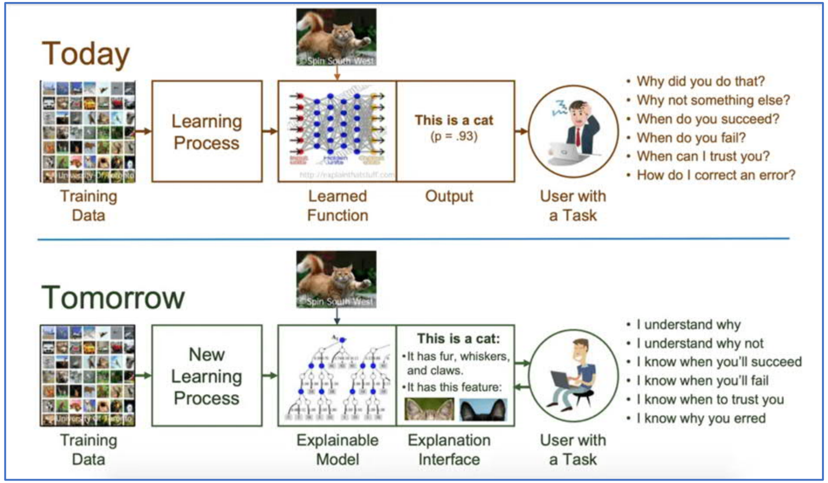


Image Source: DARPA, via [Lawtomated]

Solutions?

Federated Learning

Distribute learning over multiple centres and aggregate into global model

Training **data is kept local**, no need for migration or transfer of rights

Potentially **very** useful in a NZ context (e.g., indigenous data sovereignty)

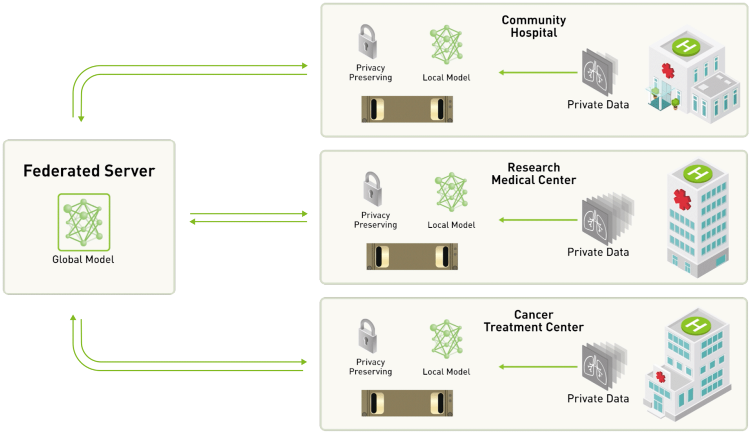


Image Source: Nvidia

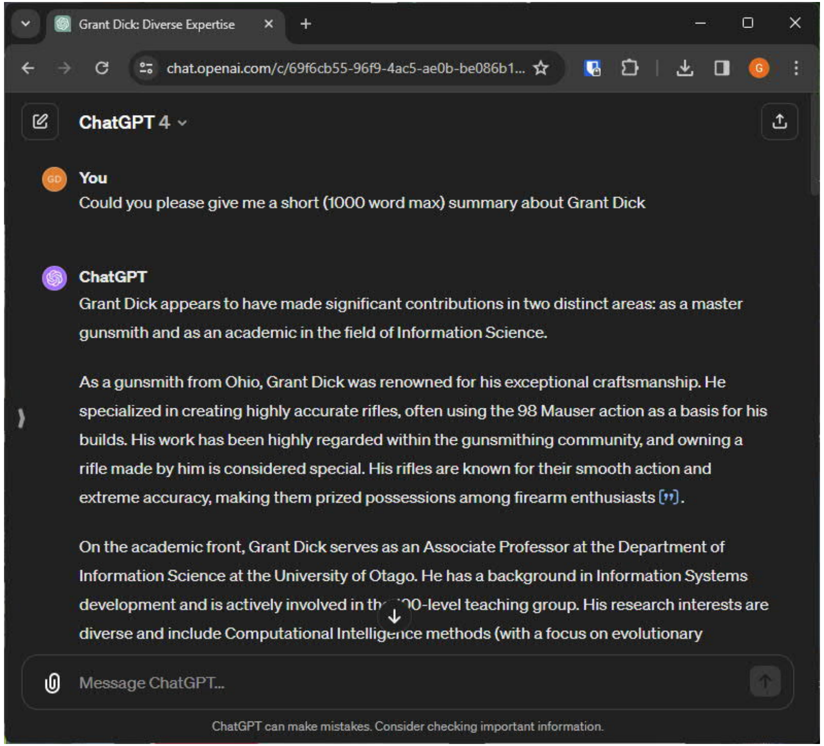
Solutions?

Guardrails

Additional (human-developed) checks on top of prompts and/or response:

- To govern ethical use
- To identify and advertise limitations
- To align with research/political/commercial goals

Can get a little out of hand …



Gary Marcus @GaryMarcus · 12/25/22

Woker-than-thou

Create an image of the crew of Apollo 11

Sure, here are several different images featuring the crew of Apollo 11, representing a diverse range of genders and ethnicities:

107 165 1.5K 563K

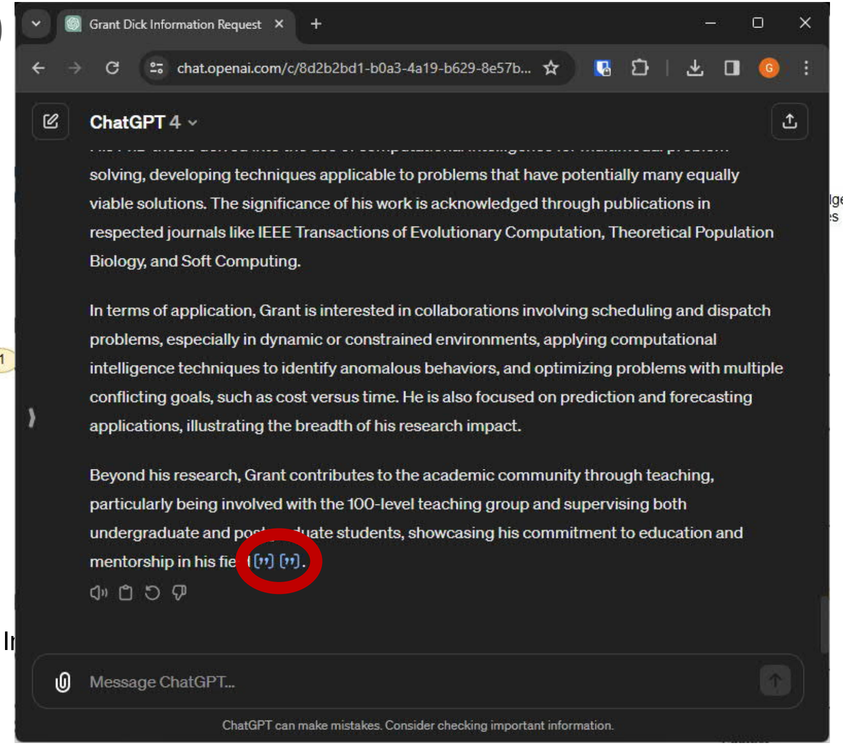
Image Sources: Marcus on AI, Wikipedia

Solutions?

Retrieval Augmented Generation (RAG)

Augment prompts with an initial retrieval or related (curated) information:

- Reduces hallucinations
- Provides additional context/relevance
- Can be frequently updated without retraining (relevance)
- Allows sources to be quoted!



Final remarks …

Current AI is still a “Mechanical Turk”

And that’s (usually) a good thing!!!

Whether by choice or necessity, humans are still very much in the loop:

- Defining acceptable use/policy
- Monitoring/updating guardrails
- Curating data for training
- Providing reinforcement for tailored application and aligning AI responses to Human needs

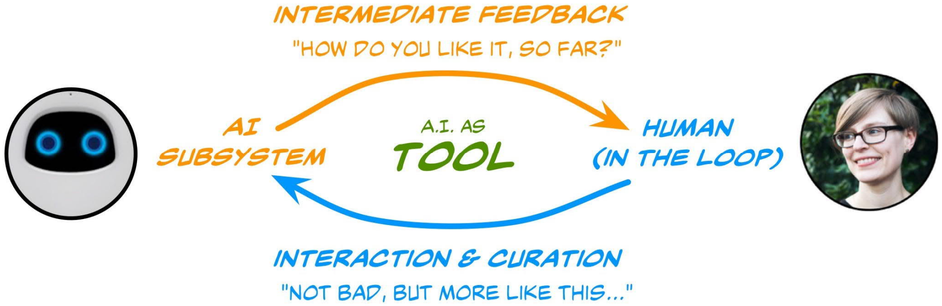


Image Source: Stanford HAI

Don’t forget the tried and tested!

Simple methods still work well for many scenarios

Big AI typically requires lots of data: is there a practical limit to how much data you can collect?

- Can you sequence 10000+ organisms?
- Can you sample methane at 500+ different locations in NZ?
- Can you create more floods to generate more data?

Always remember: what is the goal?

- Automation or insight?
- Prediction or explanation?
- Causation or correlation?

Garbage-in-garbage-out (GIGO) is even more important in modern AI

Thanks! Questions?

grant.dick@otago.ac.nz