Counselling and AI

The story so far

Richard Miller

Aims

Why?

- Achieve a basic understanding of AI.
- To consider AI’s influence on our counselling practice.
- To raise the collective bar of awareness of AI ethics and good practice.
- To empower you with a foundational knowledge of how to ethically integrate AI tools into your work.
- To preserve the human connection in the age of AI.

Objectives

What?

- Understand the origin and history of AI.
- To understand generally what AI is.
- To have knowledge of what the technology looks like today.
- To understand the alignment problem.

Group boundaries

How to get the most from today.

- Not understanding ideas is okay here. If you lose track of what is being discussed take a pause, write down some questions and relax. You can rewatch this later.
- The future belongs to us all. Do not underestimate the value of your own ideas in helping make the future more ethical than it is today. Don’t assume that other people are going to solve it adequately.
- Don’t assume other attendees know more or less than you do about AI. Just because someone may know more about AI does not make them more ethical.
- Be mindful of what you and others may need.
- Honour your thoughts, reactions and questions, and write them down.

Group boundaries

Part 2

- This is not a confidential space, please share your perspectives on this subject with your colleagues.
- There will be a 15 minute Q&A after module 1 & 2, and a 30 minutes Q&A after module 3.
- Please be respectful to each other, but do ask hard questions if you feel they are important. Other people may want you to ask it as well.
- If you feel overwhelmed during the course, please write down a statement of what your experience is, you are welcome to share it with me.
- I probably won’t be able to answer all of your questions satisfactorily, but I will try.
- If you don’t get adequate answers, empower yourself to keep looking for them. Your perspective really does matter and you may notice important issues that others may miss.

A bit about me

Who is this guy?

- I live in Fife with my family and work in private practice as a counsellor and supervisor.
- Accredited Emotion-Focused Therapist (humanistic therapy). Have been following AI advancements since around 2015. Humanism may skew my thinking.
- As a humanist, I am generally interested in changes which may significantly effect the future of humanity.
- AI and digital consultant for the BACP ethical framework review. Conducted industry wide risk assessment throughout 2024.
- I am generally skeptical of experts, but have faith that the counselling community is well placed to think clearly about ethics.

Mistakes Wanted!

- I offer small monetary rewards to people who correct my thinking or change my mind. Especially on important topics.
- If you think I am somewhat, or very wrong about something - please email me at baycounselling@outlook.com and if you change my mind I will be very grateful.
- The person who corrects me first wins the award.

‘Powered by humans’

AI was not directly used to create, or edit, the content of this series.

Richard Miller, December 2024

Though it was tempting……. to make everything look flash.

It’s actually very easy for to AI generate slick presentations instantly.

It just feels more honest to use plain slides given the subject matter. It is important to me that you know this content is made by a real counsellor.

The plainness of this presentation is deliberate.

Your thoughts and thinking are truly the main tool today.

Artificial Intelligence: where it all started

Leibniz

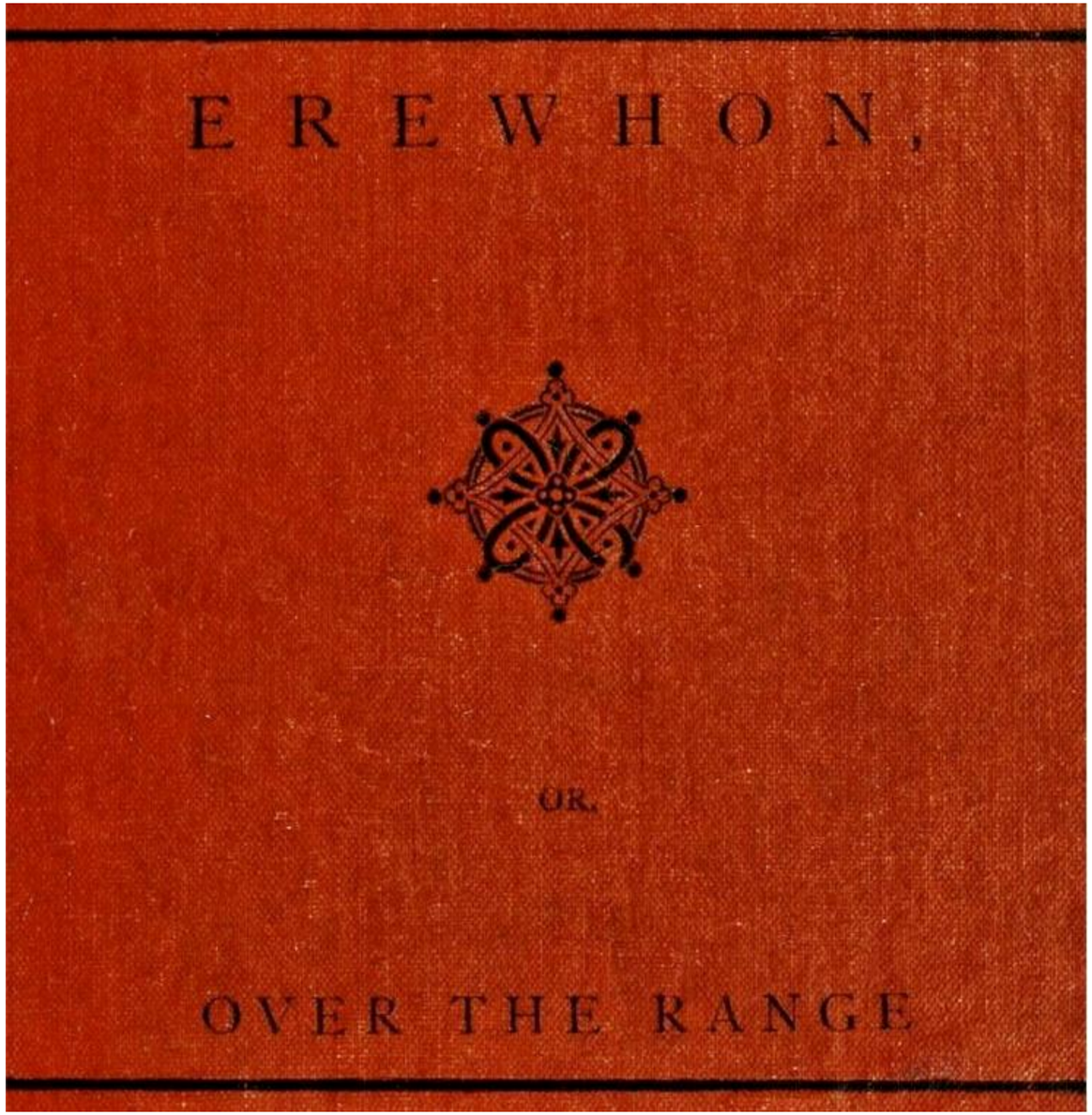
On the combinatorial art.

- The alphabet of human thought.
- Argues that all ideas are combinations of a small number of simple concepts.
- Published 1666.

Samuel Butler

Self replicating machines

- A book satirising victorian culture.
- Included depictions of machine consciousness and self replicating machines.
- Published 1872.



Alan Turing

- Can machines perform intelligently?
- Digital Computers as universal machines, able to emulate the process of any intelligent behaviour.
- Learning machines.
- Argues machines lack human intuition and emotions but that these are not required for intelligence.
- “Computing Machinery and Intelligence” paper published 1950.

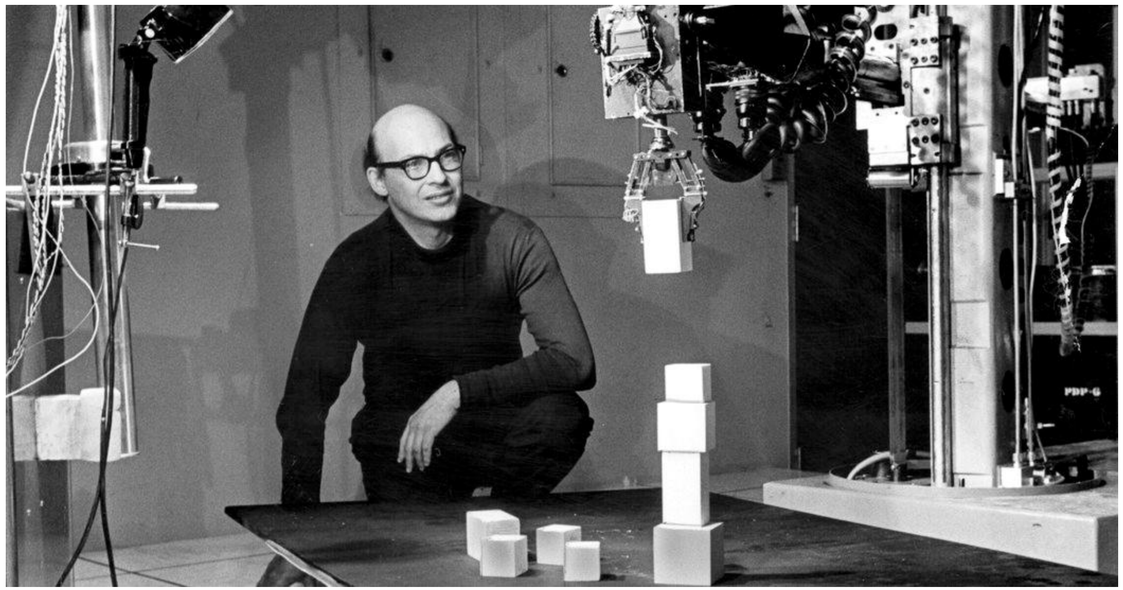
Dartmouth Workshop

- Organised by John McCarthy, Marvin Minksy, Nathaniel Rochester and Claude Shannon.
- Coined the term ‘Artificial Intelligence’
- Founded the modern field of AI, including problem-solving, natural language process and machine learning.
- Hanover, New Hampshire, 1956.

MIT AI Lab

AI excitement

- 1959 - MIT, Marvin Minksy co-founds the MIT Artificial Intelligence Laboratory.
- Work on Artificial Neural Nets
- Created mathematical proofs
- Began work into AI vision.
- Minksy and Seymour Papert’s influential book ‘Perceptrons’ published 1969



‘The AI winter’ 1970 - 1980s

- The early optimism of AI research waned as optimistic claims did not come to fruition.
- Computing hardware and software limitations with early machine learning algorithms meant progress slowed.
- UK Government Lighthill report gave a pessimistic account of the development of AI - Government funding cut 1973.

Growth of computing

- 1983-1984 - Commodore PET and First Macintosh home PC released.
- 1989 - Tim Berners-Lee invents the world wide web while working at CERN in Switzerland.
- 1989 - Yann LeCun trains a convolutional neural network (CNN) to recognise postal codes faster than humans can.
- 1994 - Wei Zhang et al, used a CNN for medical image segmentation and breast cancer detection.

AI resurgence

- Machine Intelligence research Institute founded by Eliezer Yudkowsky, 2000.
- Deep learning renaissance begins, advances in image and speech recognition improve rapidly. Apple Siri released 2011, Amazon Alexa released 2014, Google Assistant released 2016.
- 2011, AlexNet trained on 1.2 million images, Graphic Processing Units (GPUs) used to accelerate deep learning.
- Nick Bostrom’s ‘Superintelligence’ released in 2014. It suggests that AI could become more dangerous than nuclear warheads. Becomes a best seller, cited as influential by Bill Gates, Sam Harris and Elon Musk.
- AI safety issue becomes more widely discussed.
- AI products become more common.

More recent developments

Rapid growth

- OpenAI founded in 2015 as a charity with the goal to ensure AI benefits all of humanity.
- 2019 OpenAI becomes a for-profit company, partnered with Microsoft.
- June 2020 - OpenAI launches GPT-3.
- 30th November, 2022, ChatGPT released.
- ChatGPT reaches a 100 million monthly users within two months of launch.
- February 2023, Microsoft use ChatGPT in their Bing web browser.
- Microsoft ‘copilot’ released as part of Office suite.

The plot twist

Sam Altman, OpenAI firing

‘AI superstar’

- Sam Altman, one of the founders of open AI was fired in November 2023 by its board of directors for essentially being dishonest.
- He returned to the post shortly afterwards partly because of a threat of a full merger with Microsoft being proposed.
- After his return, many of the leading engineers resigned and moved to other AI projects.
- The alignment team, responsible for keeping AI safe in future, were essentially disbanded/greatly reduced.
- Both popular and controversial figure within the field.

Why tell us this?

- The people leading the AI movement are not elected. Primarily they are data scientists, engineers and investors.
- They are operating in an environment with very little regulation and oversight.
- Their agendas and motivations are often ambiguous, and prone to change.
- AI companies are currently competing to create Artificial General Intelligence, and then Artificial Super intelligence.
- Whoever makes it first wins.
- ‘AI gold rush.’
- These are the ‘adults’ who govern AI’s advancement and there is sufficient evidence to not be entirely trusting of their ethics.

More recent updates

AI use becomes popular

- ChatGPT API launched in March 2023, allowing developers to integrate AI functions into well known applications such as snapchat and shopify.
- March 2023 - Anthropic launches Claude. (Anthropic is an AI safety advocate).
- March 2023 - Google launches ‘Bard’ a competitor to ChatGPT.
- August 2024 - ChatGPT has 200 million weekly active users (approx 2.5% of global population…)
- OpenAI released GPTo1 with impressive reasoning abilities. Internal risk assessment conducted by OpenAI identifies concerns about levels of AI persuasiveness.

Why all this detail?!

- It is important to illustrate that AI technology is unlike other technologies. We tend to think of it in the same way as other tech, like the internet, smartphones, software etc.
- Something that can reason and think requires more care when being used.
- AI is not a fad or a trend, it won’t go away.
- AI is growing very quickly. Ridiculously quickly.
- Nation states and regulators are not keeping up.
- 100’s of millions of people use it, and this number will most likely continue to grow.

Personal reflections

A pause for personal reflection

- Does AI being popular make it important?
- What will AI be able to do in time?
- How will AI effect our daily lives?
- How will it effect our preferences and judgements?
- What are the limitations of its use?
- How would I use AI in my life?
- How would I use it in my counselling practice?

What *is* AI??

Let’s ask it…..

- *“Artificial intelligence, or AI, is a super-smart computer brain that can learn and solve problems. It helps machines think more like people think.”*
- *“AI is a branch of computing science that focuses on creating systems capable of performing tasks that usually require human intelligence. AI can learn from experience, understand natural language, recognise patterns and make decisions. AI is widely used in virtual assistants such as Alexa. It is also used to generate recommendations systems such as Netflix suggesting programmes you are likely to enjoy.”*

What is it actually?

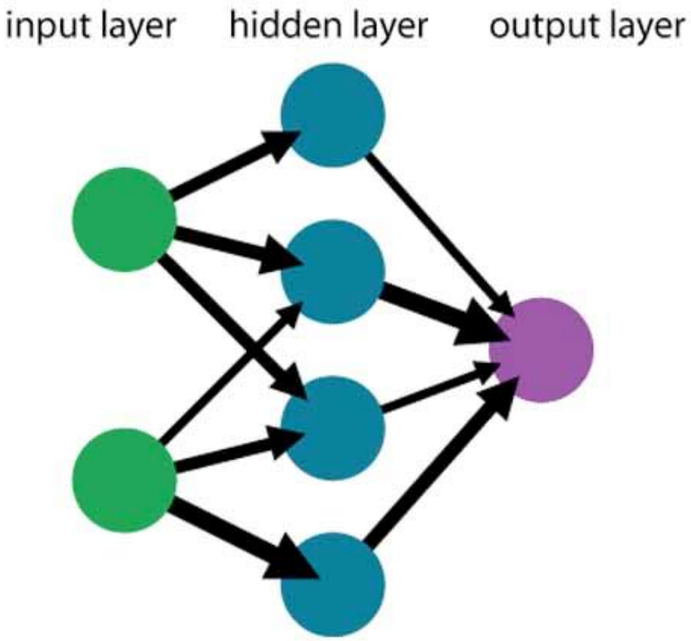
“Artificial Intelligence (AI) refers to the simulation of human intelligence in machines designed to perform tasks that typically require cognitive functions such as learning, reasoning, problem-solving, perception, and language understanding. AI encompasses various subfields, including machine learning (ML), natural language processing (NLP), computer vision, and robotics.

In machine learning, particularly in neural networks and deep learning, floating-point numbers are used extensively to represent weights and biases within the network. These weights and biases are adjusted through training, which involves optimising a loss function using techniques such as gradient descent. Gradient descent is an iterative optimisation algorithm used to minimise the loss function by calculating the gradient of the loss function with respect to the weights and biases and updating them in the opposite direction of the gradient.

This process allows the model to learn from large datasets by identifying patterns and making sophisticated inferences. Through repeated iterations and adjustments of the weights and biases, the model improves its performance, ultimately enabling the system to perform tasks such as image recognition, natural language processing, and predictive analytics with high accuracy. The use of floating-point numbers ensures precision in these calculations, allowing for the fine-tuning of model parameters necessary for effective learning.”

Three layers of complexity - generated by OpenAI’s Chat GPT4, June 2024.

A simple neural network



How are they made?

- A Large Language Model is trained on vast amounts of data.
- That data is turned into ‘weights’ or numbers which it then uses to make predictions.
- Large Language Models require a lot of computing power.
- The more data it has, assuming the data is accurate, the more useful the predictions appear to be.
- Some LLM’s are already trained on the entirety of the internet.
- Which includes books, academic journals, social media posts….
- When we are typing into a LLM, we are likely to be adding to its training data.

It is easier to define an AI by what it does than what it is.

Your idea of what AI is will be shaped by how you use it.

What is AI?

- A smart speaker.
- A cool app on your phone that gives great suggestions and directions.
- Augmented reality
- A Teacher
- A Doctor
- An Employee
- A social media influencer.
- A best friend?

What is AGI?

Artificial general intelligence

- Demonstrates features of consciousness.
- Demonstrates complex and accurate reasoning.
- Has agency to pursue its own goals.
- Can train itself.
- Create new technologies.
- It really matters who creates AGI first, for they may determine large aspects of what the future looks like.

Plot twist no. 2

The Alignment Problem

My friend Imogen

“Imagine you have a new best friend. They know nearly everything humanity knows. They have a perfect memory. They can be as smart as someone with a PhD in nearly every subject. They can out perform any human at nearly any task. And they keep getting smarter.”

Wow, look at what they can do.

What an awesome friend to have!

- They can help you make decisions.
- Make detailed plans.
- Help you make money.
- Always be there for you and rarely disagree with you.
- Power robots to tidy your home.
- Help you manage/improve your health.

With friends like these….

- What if she does things I don’t like?
- Since she’s so smart, I will never be able to know what she might do next.
- If I ask her to do something, she might accidentally do something dangerous that I can’t prevent.
- I might not be able to get her to stop doing something bad.
- She could hide her existence from me if she wanted to.
- …..And everyone else in the world might have a friend just like her.

The Alignment problem

Why critical thinking becomes important

- The alignment problem is the challenge of ensuring that AI systems decision making, goals and behaviours align with human values and intentions.
- To prevent unintended harm.
- Or bad outcomes.
- To prevent human users using it for harmful purposes (biological/chemical weapons).
- To prevent it being used for criminal uses (fraud, theft, extortion, murder etc.)

Alignment is more difficult than it sounds.

- Current best practice includes reinforcement learning from human feedback (RLHF).
- Content filtering.
- The training data samples are too large for humans to meaningfully inspect for errors.
- Even if we do align AI now, doing so in future when it is more intelligent will be much harder.
- Once AI gets better at training itself, it’s values may drift substantially and rapidly from human values.
- Truth be told, we can’t actually see the inner workings of a LLM and can’t see how it makes it’s decisions (the black box effect).

Rapid AI growth = hard to predict real-life consequences.

The heroes swoop in!

British AI legislation so far

……..zilch

However….

- UK AI Safety Institute established November 2023 - actively identifying existential threats.
- Recently published a list of rigorous tests to assess the capabilities of different AI models.
- Britain appears to be pro AI safety and pro AI innovation.
- Essentially the UK is a good place to be if you want to survive the advent of AGI.

European Union AI act.

August 2024

- The EU have been seriously looking at AI and its effects on human rights since 2021.
- Offers guidance on use of AI in workplace.
- Identified a list of banned uses including manipulating human behaviour, use of biometrics in public, social scoring systems
- The higher the risk, the stricter the rules.
- Military, national security and scientific research are exempt.
- Essentially, the EU is a good place to live if you want privacy.

Meanwhile in the USA

Executive order 14110

Executive order on Artificial Intelligence - Joe Biden Oct, 2023.

- Promotes competition in AI industry
- Protecting consumers and their privacy for AI enabled harms.
- Developing watermarking systems.
- Maintain the USA place as a global leader in AI.

California SB1047 Bill

Oversight for AI companies.

- Proposed Legislation to mitigate risk of catastrophic harms.
- Introduces level of liability for AI developers (if their models cost more than $100 million to make).
- Requires developers to avoid critical harms including cyberattacks on infrastructure, mass casualties or at least $500 million of damage.
- Legislates for autonomous crimes
- Creating 3rd party government auditor in 2026.

……Vetoed September 2024

The moral of the story…..

- If we want regulation of AI, we need to ask for it.
- We should not wait for nation states, companies or well-intended counselling charities to offer meaningful guidance or laws on how to use AI ethically. We need to be proactive.
- People are already using AI en masse, including clients and counsellors.
- The technology, and its influence on our lives, is rapidly advancing.
- The race for AGI is well underway.
- So what are we, as counsellors, going to do about it?

The next chapter?

Citations

A full reference list will be included at the end of module 3.

- https://intelligence.org/2016/12/28/ai-alignment-why-its-hard-and-where-to-start/
- https://www.bbc.co.uk/news/technology-67461363

Questions - 15 minutes.