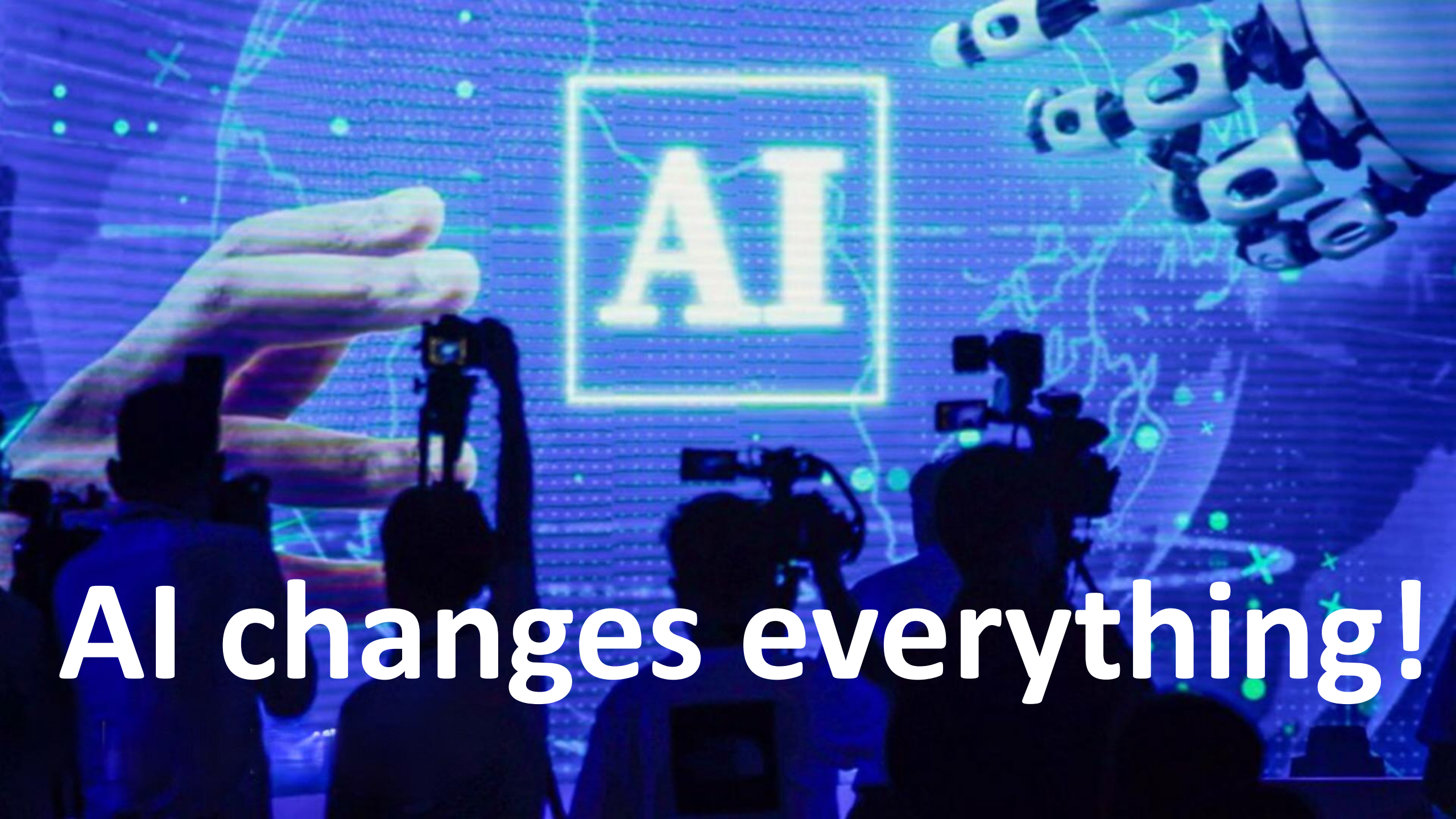


AI changes everything!

1980

1985 - 2005

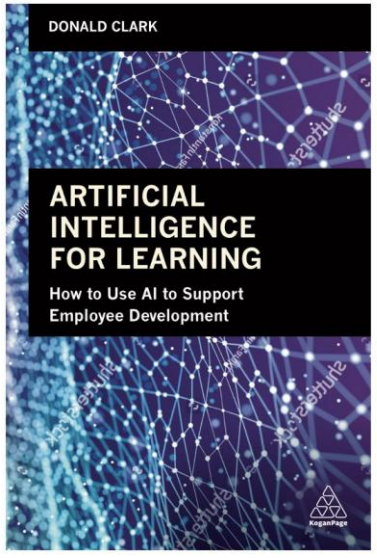
2013-21

CogBooks

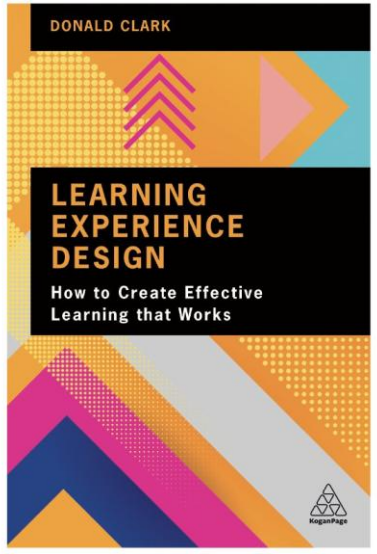
2006-21

2014-

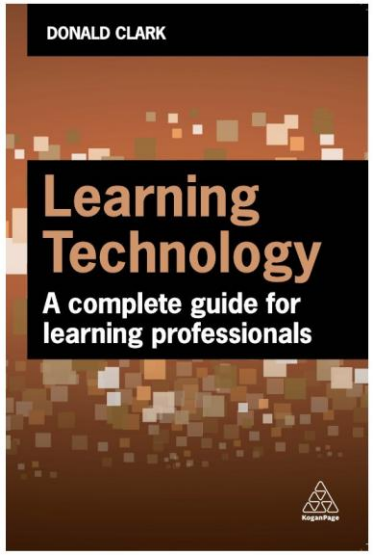
2021



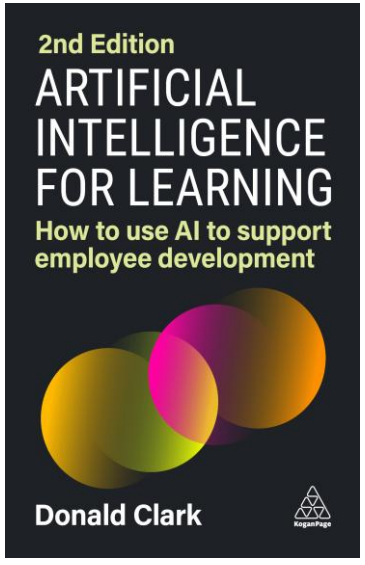
2022



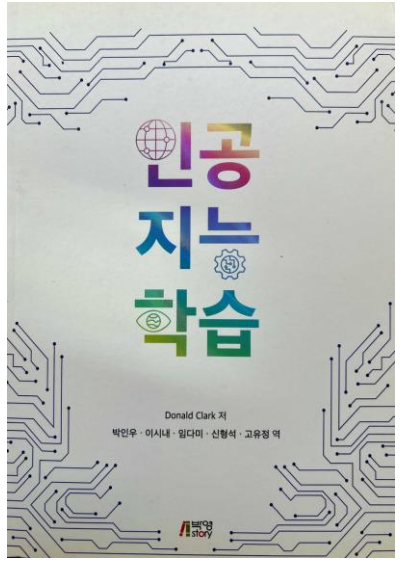
2023

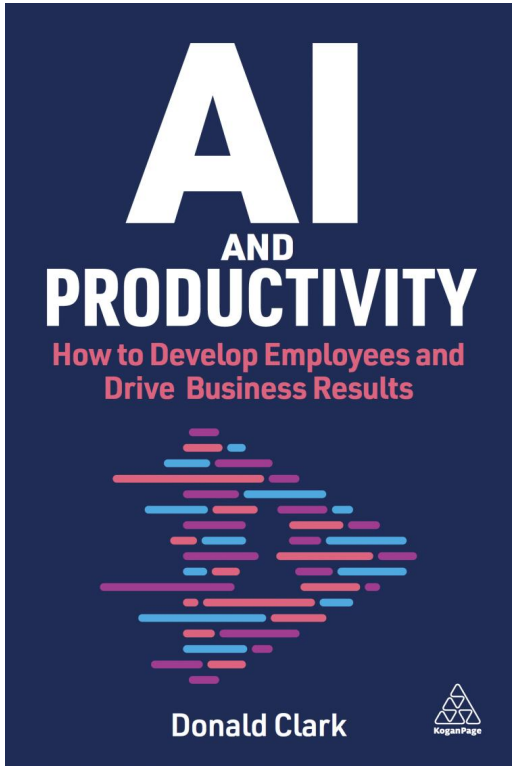


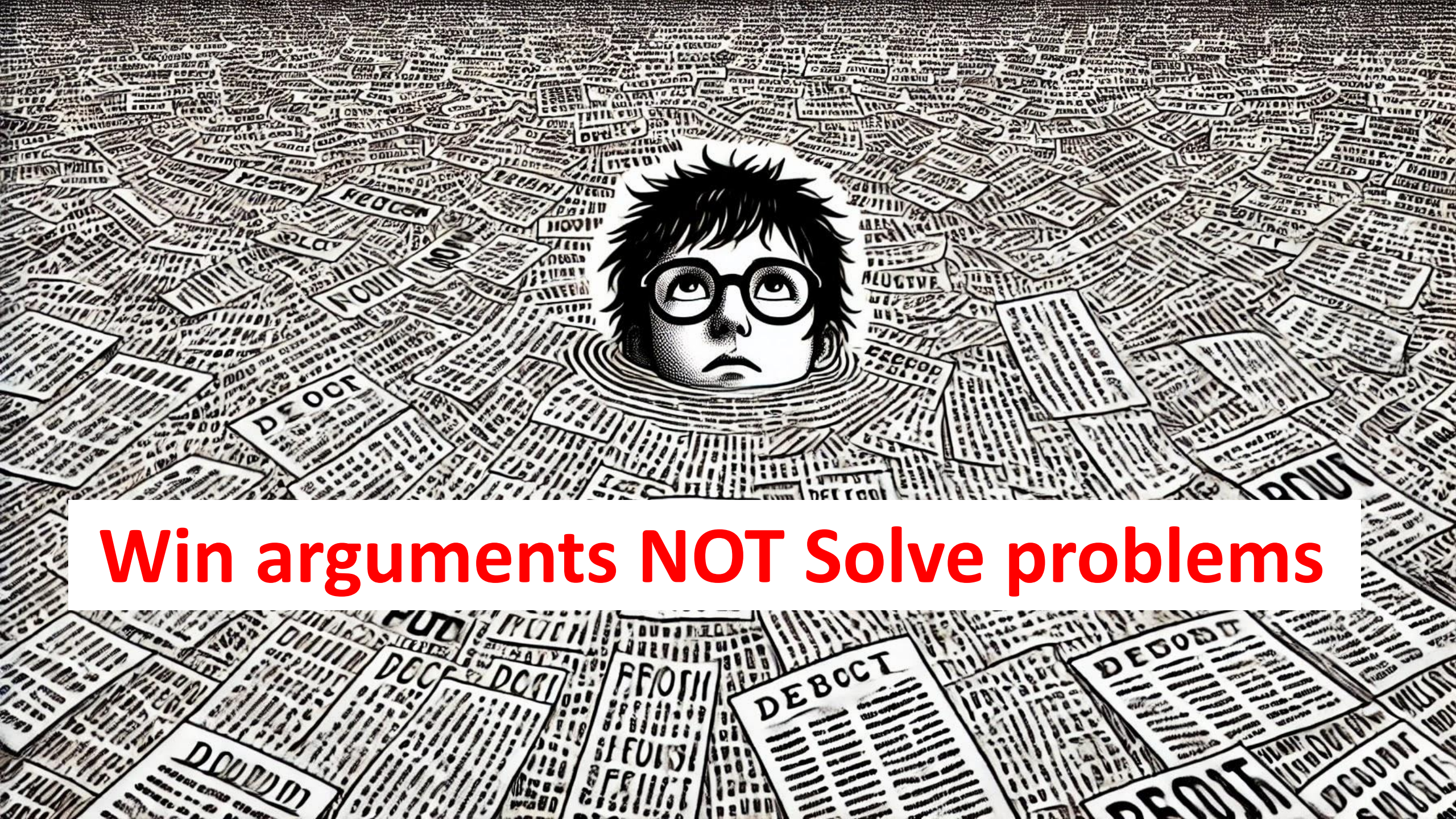
2024



2025






Win arguments NOT Solve problems


SKILLS

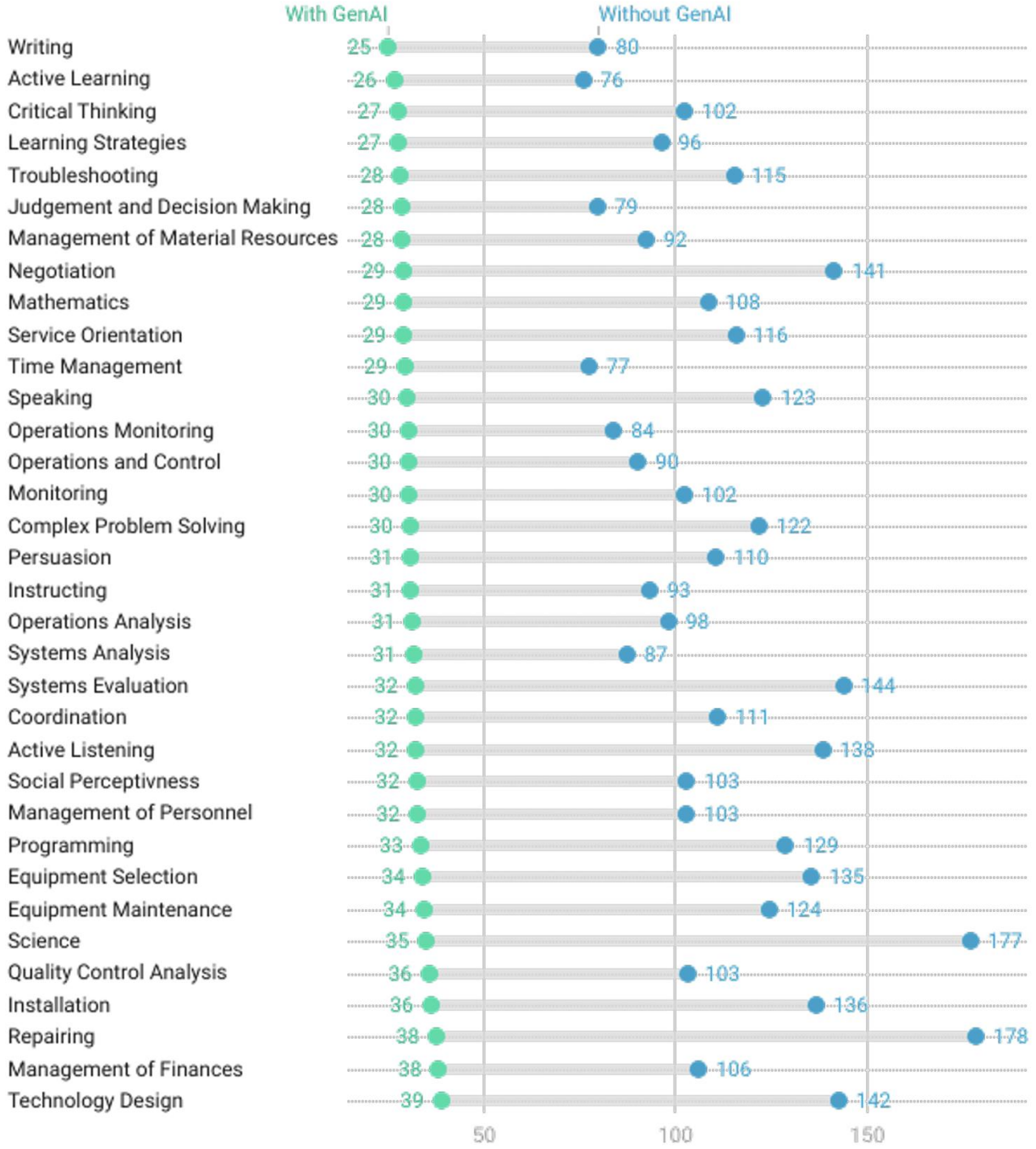
With GenAI

Without GenAI

LARGE time decreases

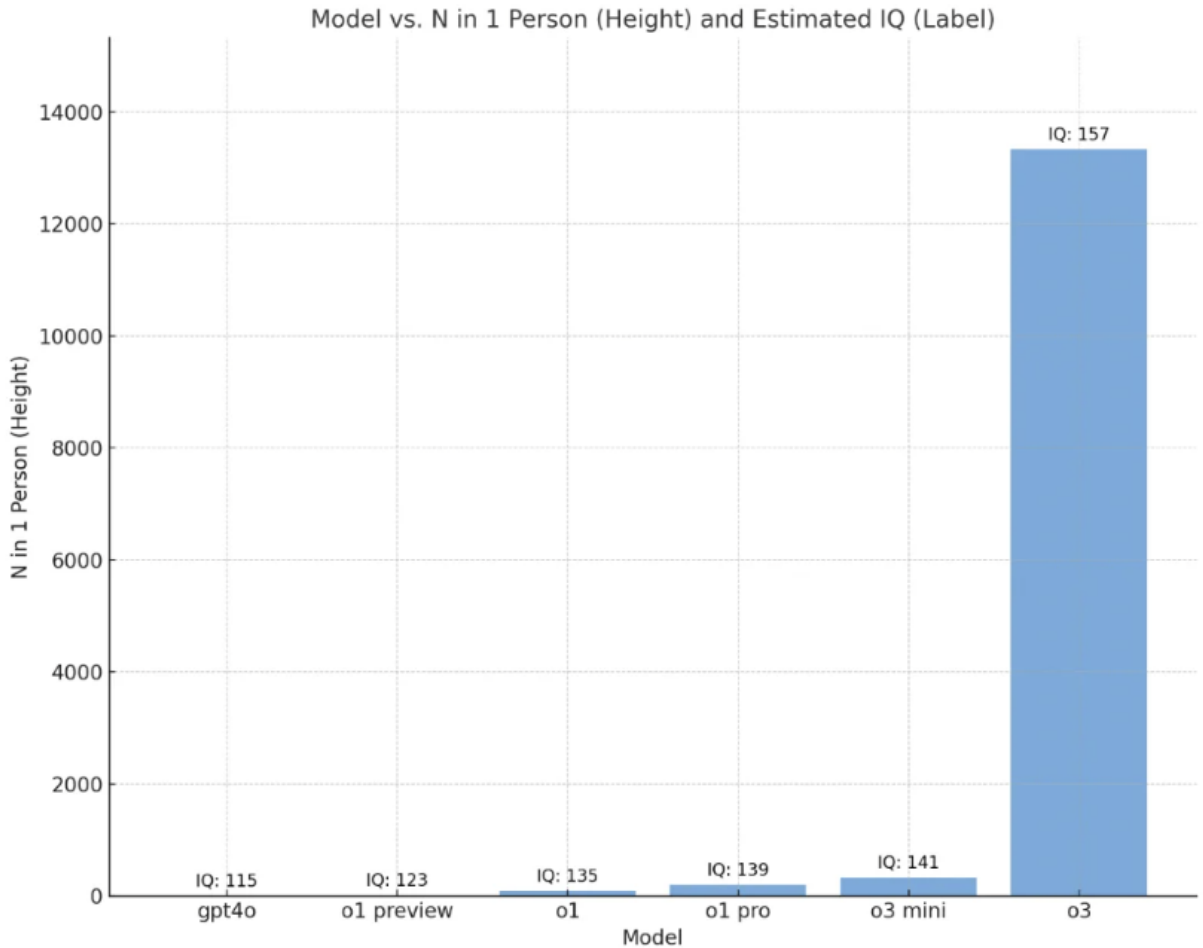
69 -80%

Figure 11: Average number of minutes to complete a task with and without Generative AI



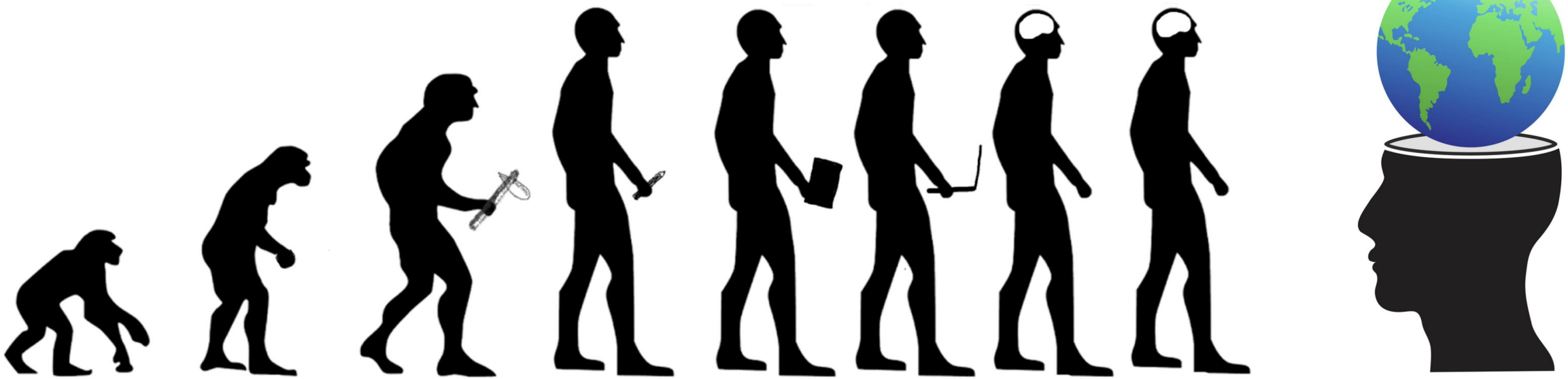
Hartley Feb 2025 The Labor Market Effects of Generative Artificial Intelligence

AI IQs skyrocketed in 2024



~1 in 6 people have an IQ as high as GPT-4o

Only ~1 in 13,333 people have an IQ as high as o3



Bipedal
5,000,000

Marks
500,000

Cave art
50,000

Writing
5,000

Printing
500

Computers
50

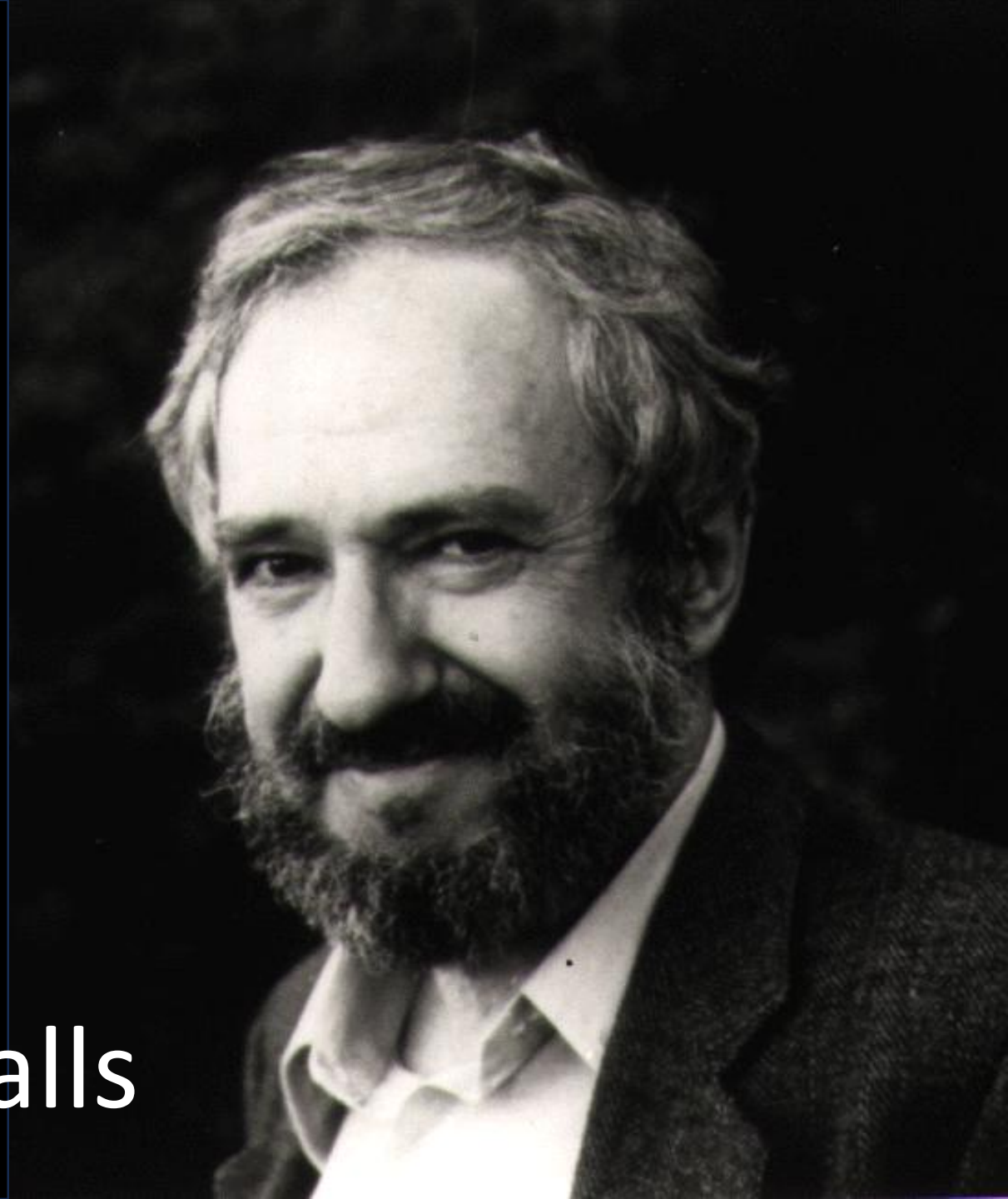
AI
5

AI
5 days

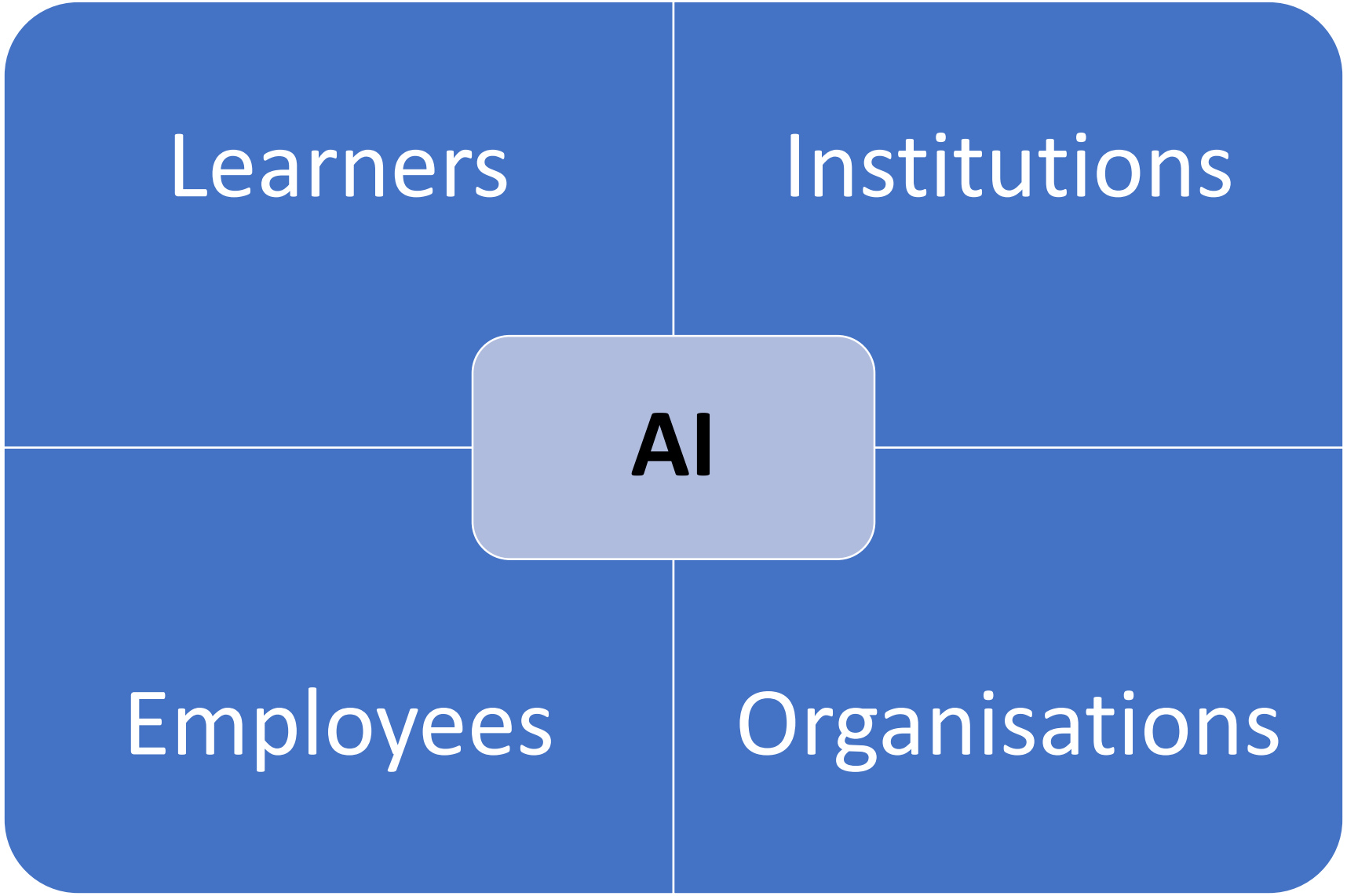
170,000 yrs to read 8 hrs/day

HIGH Ceiling

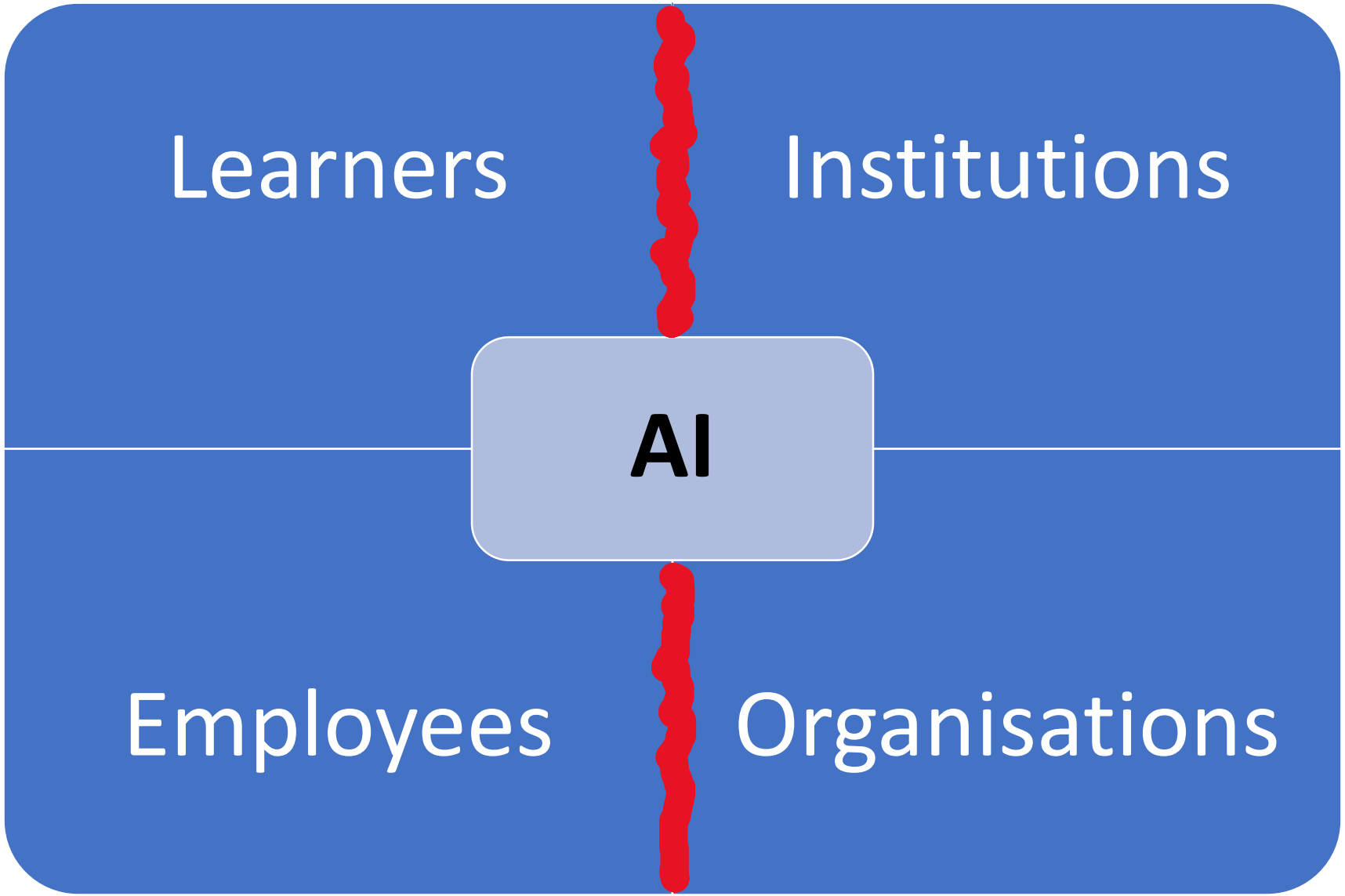
LOW Floor **WIDE** Walls



AI is at a crossroads…



AI is at a crossroads…



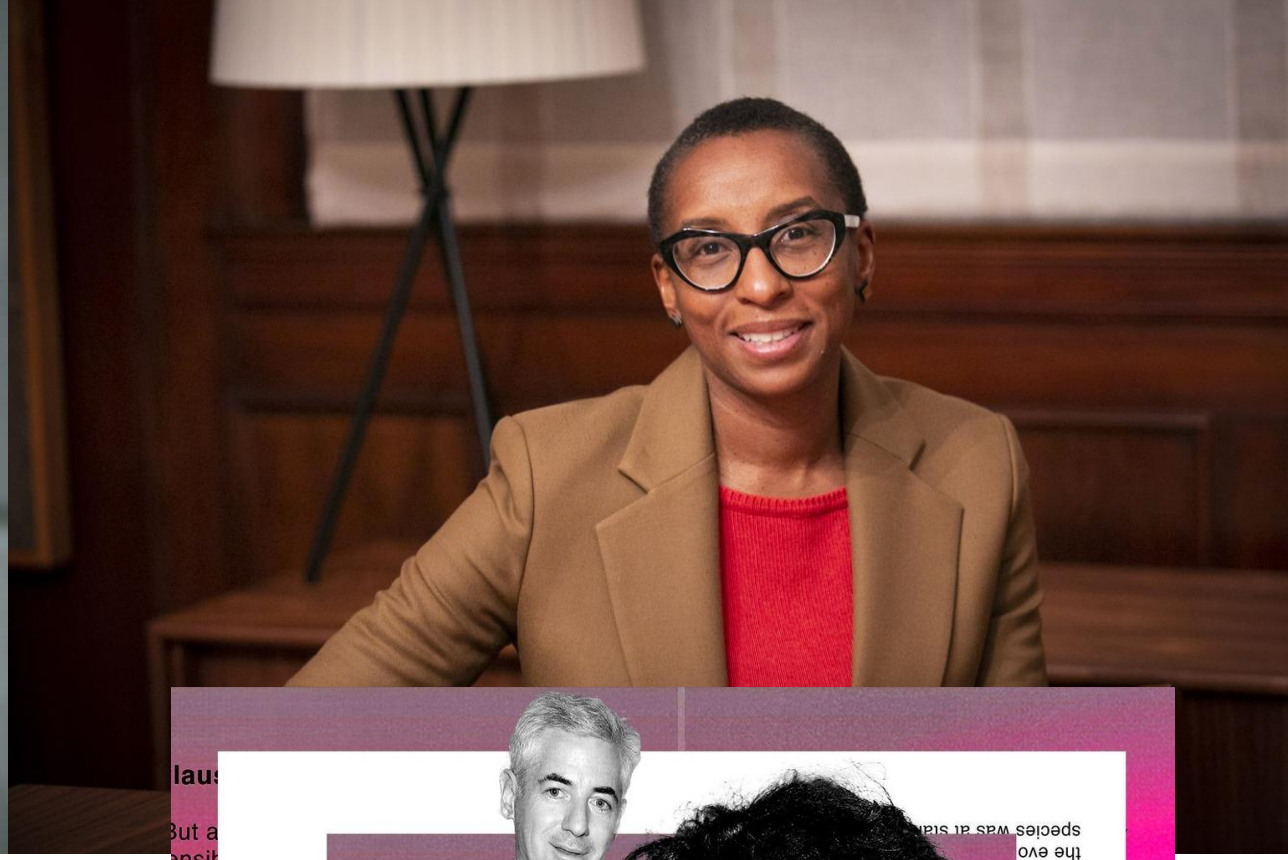
AI on the SLY?

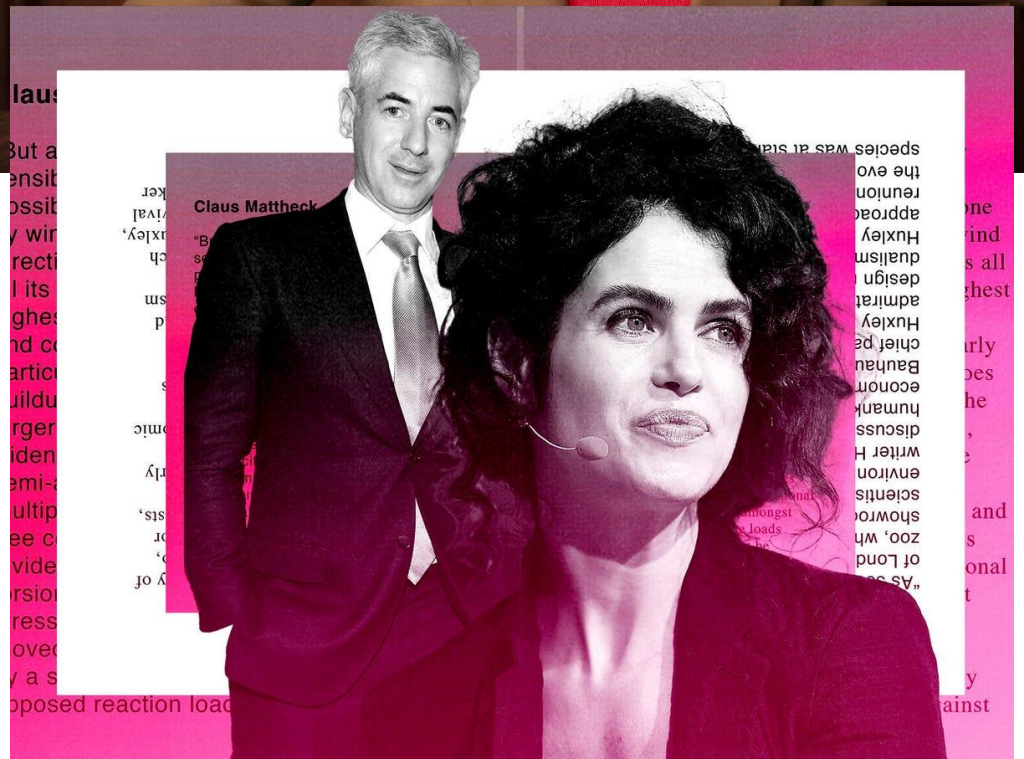
Widespread **unsanctioned** use of AI tools to save time

78% used AI tools **NOT** provided by organisation

95% in Higher Education

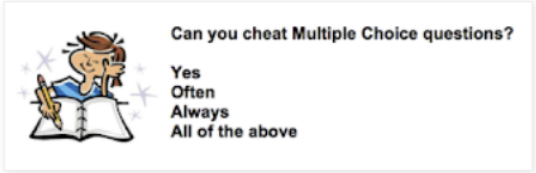
https://www.shrm.org/executive-network/insights/people-strategy/year-ai-hr-ibm-lamoreaux-spring-2024




Claus Mattheck

All of the above: 20 ways to cheat MCQs

All of the Above – 20 ways to cheat MULTIPLE CHOICE questions



Many multiple choice questions are poorly written. What better way to expose these errors than write a cheat-sheet for learners?

Of course, writing good test items is far more difficult than many imagine. Many make obvious mistakes. An interesting way to coming at this problem is to do some reverse engineering. If you think this doesn't work, think again. Poundstone number crunched 100 tests with a total of 2456 questions to get some of these statistical biases.

Second-guessing the test designer
So here goes with my 20 ways to cheat Multiple Choice tests:

1. **Skip the hard questions**, mark them with a cross, and go back to them. This means you'll not lose marks for unanswered easy questions.

2. **Cover the options and try to answer**. Prevents being misled by clever wrong options.

3. **If in doubt choose 'B'**, poor questions designers do not truly randomise the right options and have a bias towards 'B'. Next best is 'C'.

4. **If in doubt choose the 'longest option'**. Question designers often cannot make a right option any shorter, but have complete freedom with wrong options. This is quite common.

5. **If in doubt choose TRUE**, in true/false questions, as they come easier to mind for designers.

6. **Reverse answers**. Statistically, there is more T/F alteration in tests than in truly randomized sequences, as the brain struggles to randomise properly. So, if you're sure you've got one right, reverse the next answer.

7. **Eliminate the outlier.** Look for similarities in options and eliminate outliers e.g. 4p-q, 2p+q, 4p+q, 3p+q. Look for these internal patterns.

9. **Favour options with careful qualifiers**, such as 'sometimes, occasionally etc.' as tested knowledge usually has more finite than absolute qualities.

10. **Be wary of options with absolute qualifiers**, such as 'always, never etc'. As these are often too definite to be correct.

11. **Choose a middle order option** i.e. out of 100, 150. 200, 250, choose 150 or 200. Designers tend to have a bias, where right answers tend to be lower than the highest and higher than the lowest option.

12. **For questions that demand an 'except' or 'not',** mark each option with a T for true and F for false against each option. And underline the word 'not' as it's sometimes missed.

13. **'All of the above' and 'None of the above'** are both significantly likely to be correct. For it to be correct, the writer has to design options that were all correct, so, if you can't spot any wrong answers, or see that two or more are correct, it increases the probability of 'All of the above' being correct. Similarly with 'None of the above'.

14. **Typo or punctuation error**, the option is likely to be wrong. Writers tend to proofread correct answers only.

15. **Look for grammatical agreement** between the question and its options; 'An.....' and words starting with vowels or agreement between subject, object or verb.

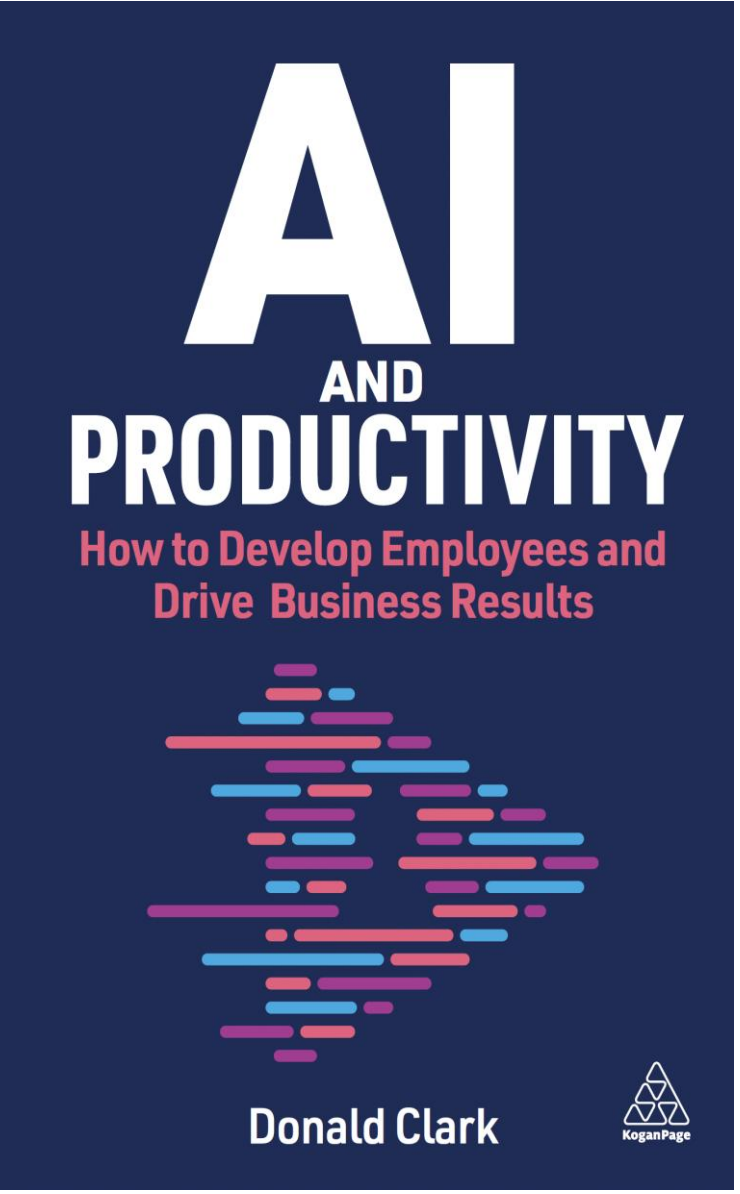
16. **If you're stuck, go with the 'Least bad rule'**. Eliminate least likely answers first.

17. **Look for clues about answers from other questions**. Designers often, unintentionally, put clues, even answers, to questions in other questions.

18. **Ignore never heard of answers**. If you've never heard of the answer, it's likely to be made up and incorrect.

19. **Go with your first impression**. The more you read, the more you tend to read into the wrong options.

20. **Always guess**, unless there is a penalty. It's a 1 in 4 chance, so don't give it up.



BRAINS, AI, AGENTS & AGI

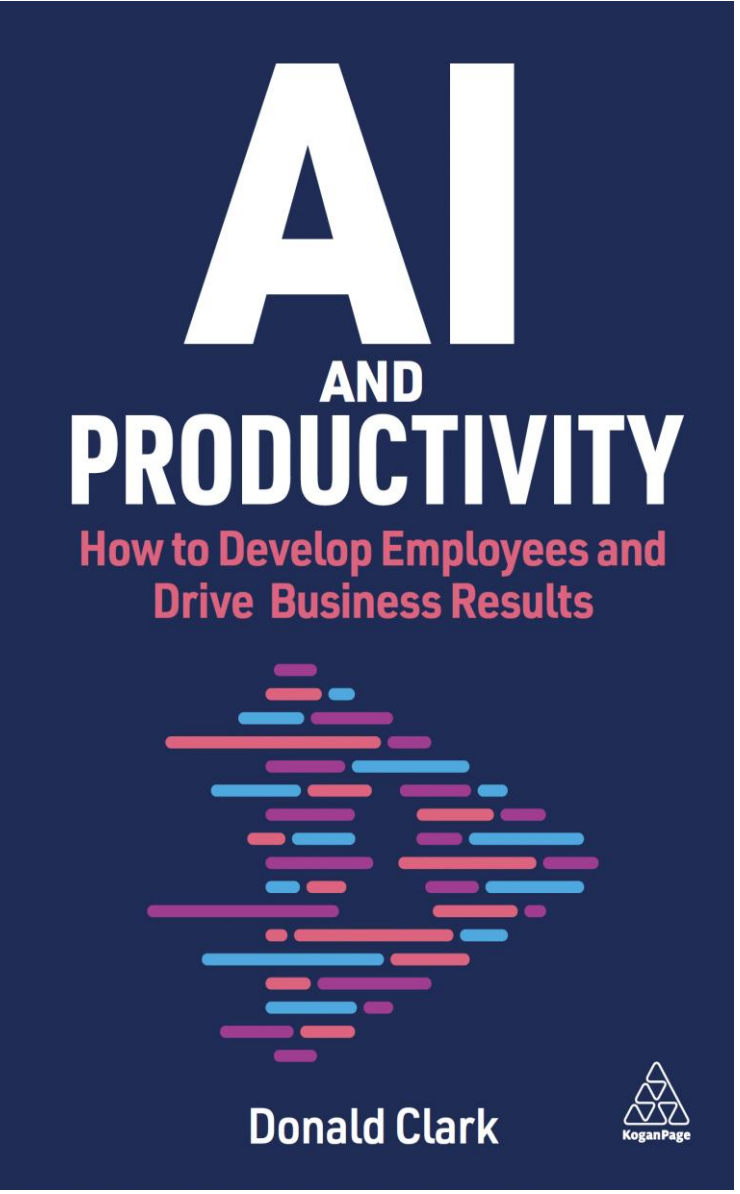
Cognitively capped

20+ years to educate
Forgets
Cognitive overload
Fallible memory
Cognitive biases
Sleeps 8 hours
Can't upload
Can't network
Doesn't scale
Dies!

Made rocks think

Memory, Reasoning, Research, Agents

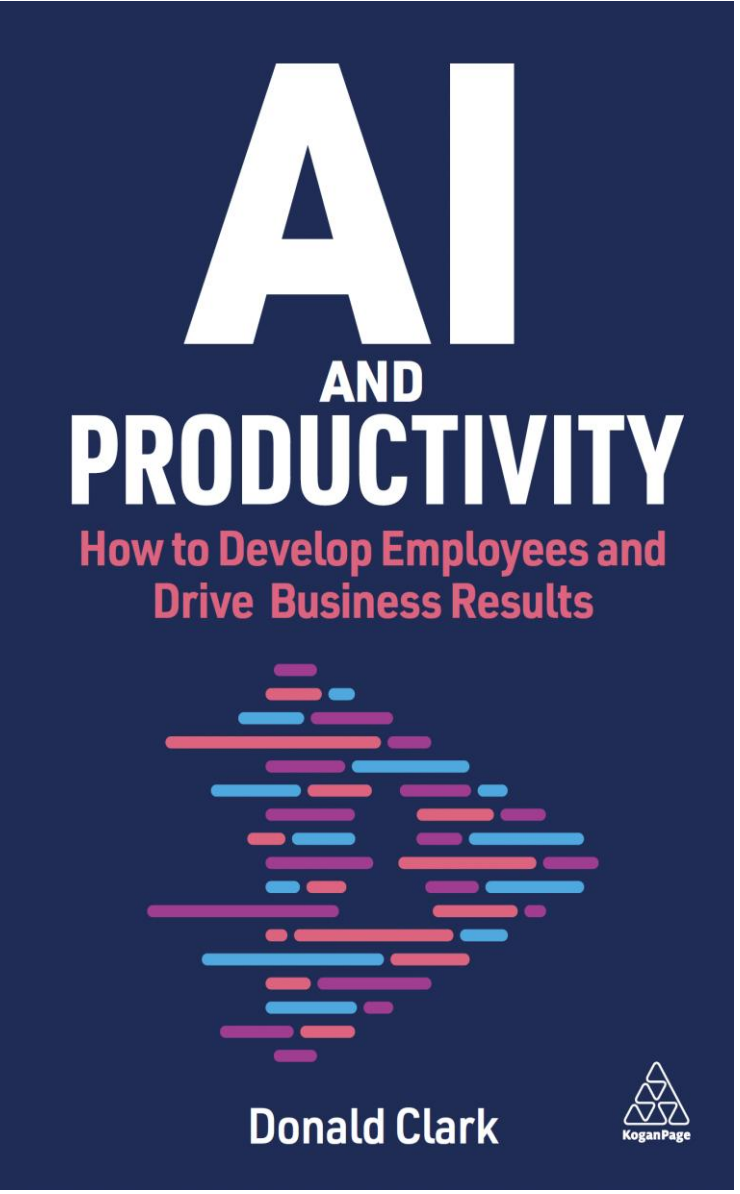
AGI



EVIDENCE of productivity?

AI boosts productivity but unevenly
Academic, field research, reports, surveys

Jagged **frontier**



Productivity **PARADOXES**

Behavioral paradoxes

Bias paradoxes

- **confirmation**
- **anthropomorphic**
- **status quo**

Procrastination paradox

Organisational paradoxes

Perfection paradox

Busyness paradox

Parkinson paradox

Boiled frog paradox

Technological paradoxes

Solow paradox

Generality paradox

Legacy paradox

Moravec paradox

Economic paradoxes

Jevon

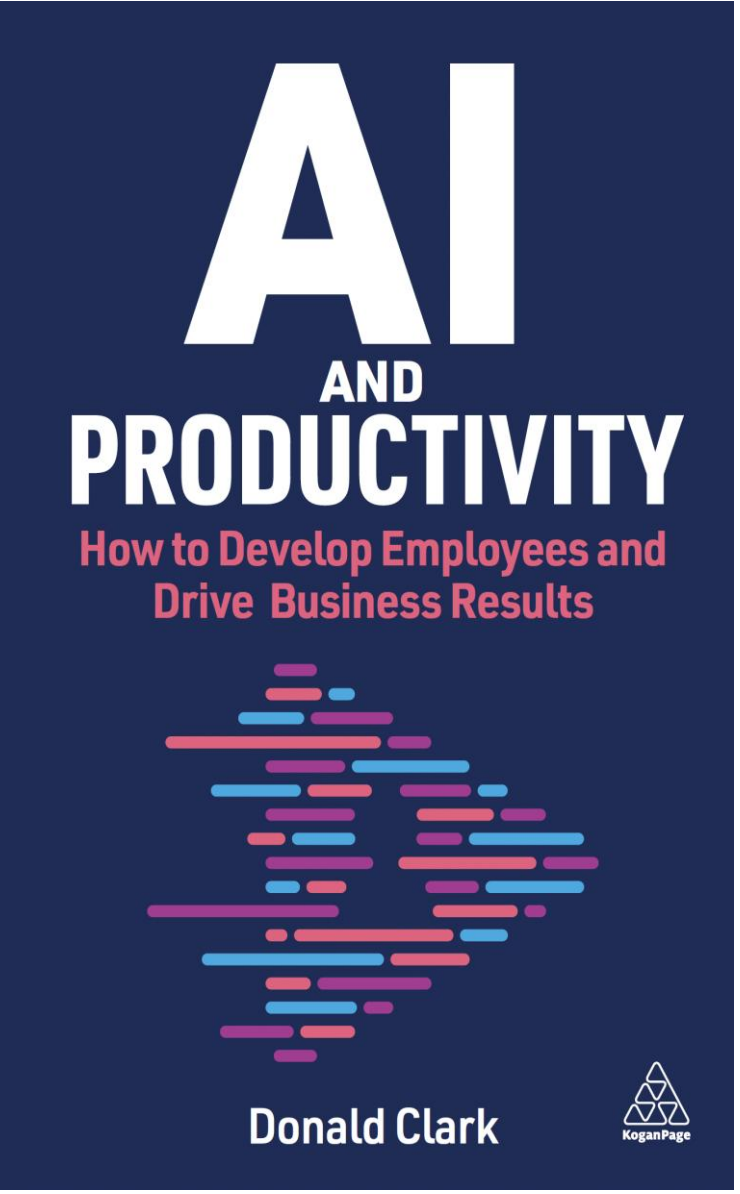
Spoons

Easterlin

Turchin

Pollyana

Empowerment



HUMANS-IN-THE-LOOP

Humans-taken-out-of-the-loop

Elevated productivity

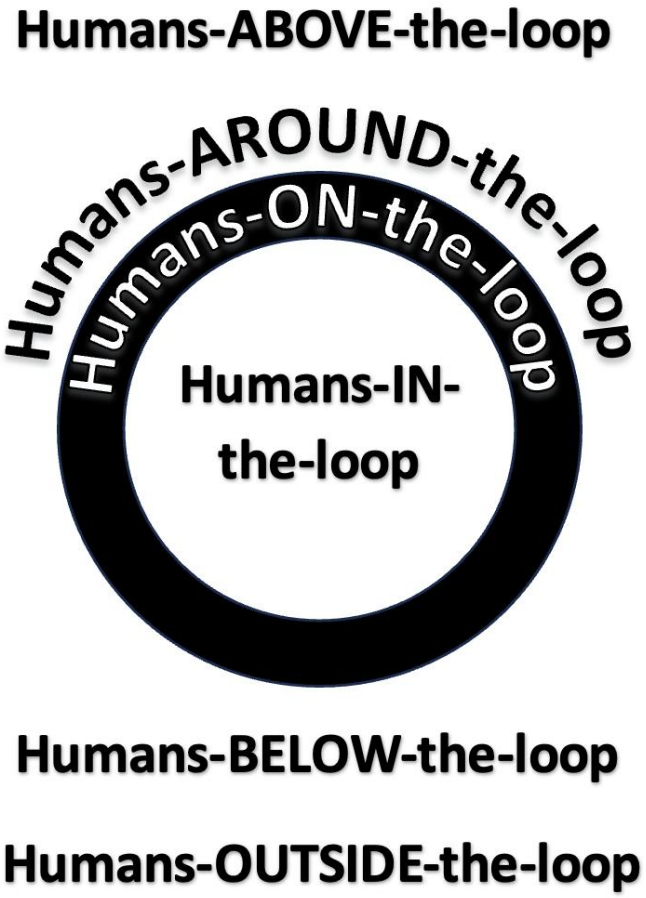
Centaurs and cyborgs

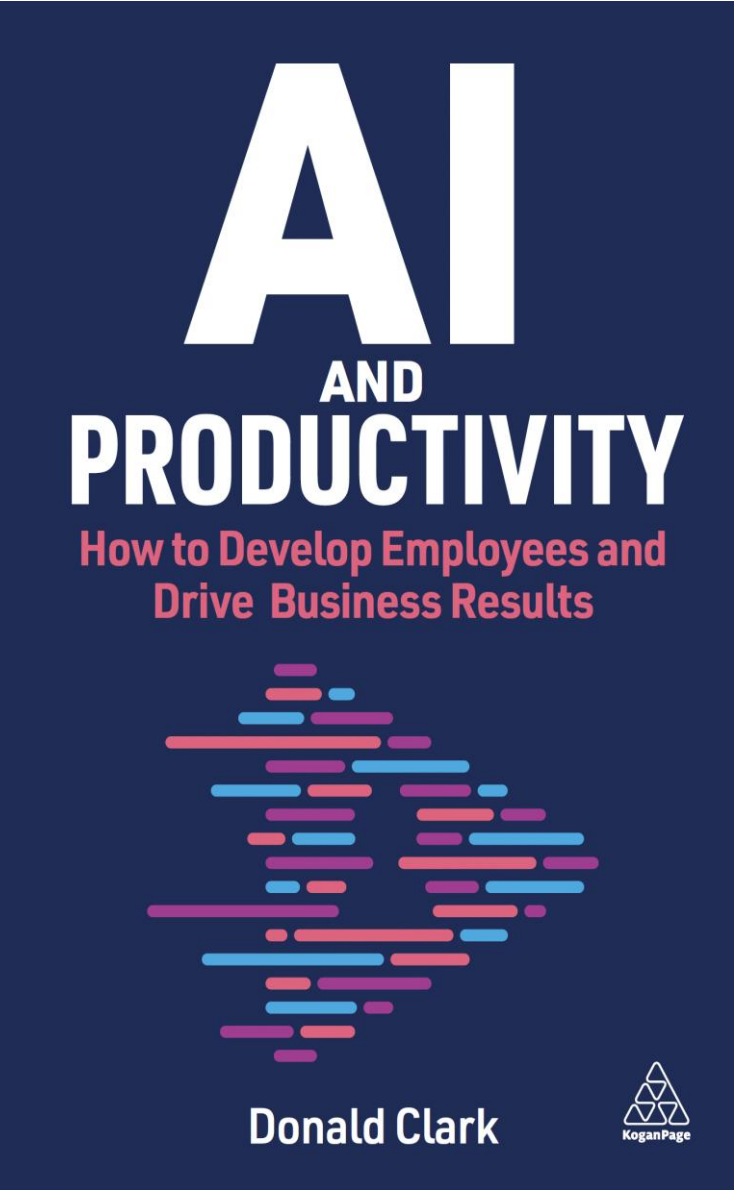
Experts-in-the-loop

Agents-in-the-loop

Robots in the loop

Six GENERAL Levels of Autonomy



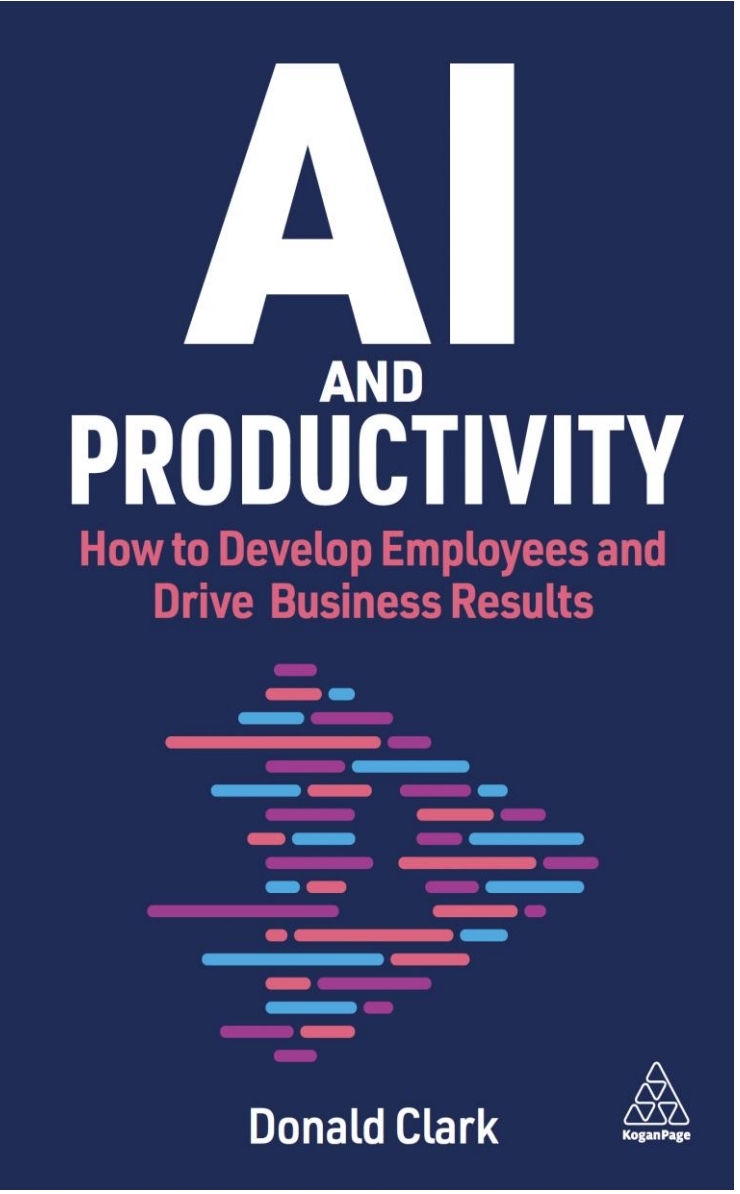


MULTIMODALITY

Productivity and accessibility

Multilingual

Text
Audio
Images
Avatars
Video
Art

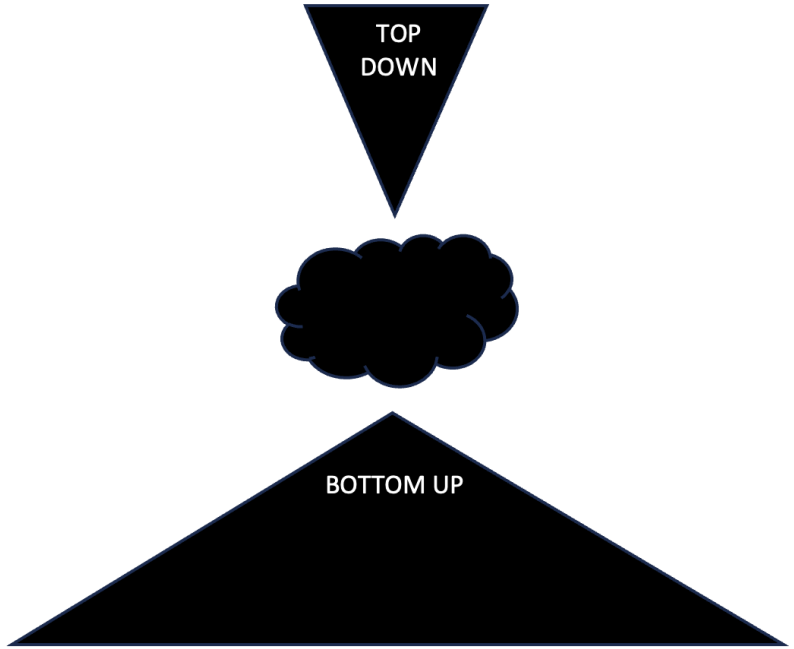


DEPLOYMENT in organisations

Inconvenient truths

Get the ball rolling

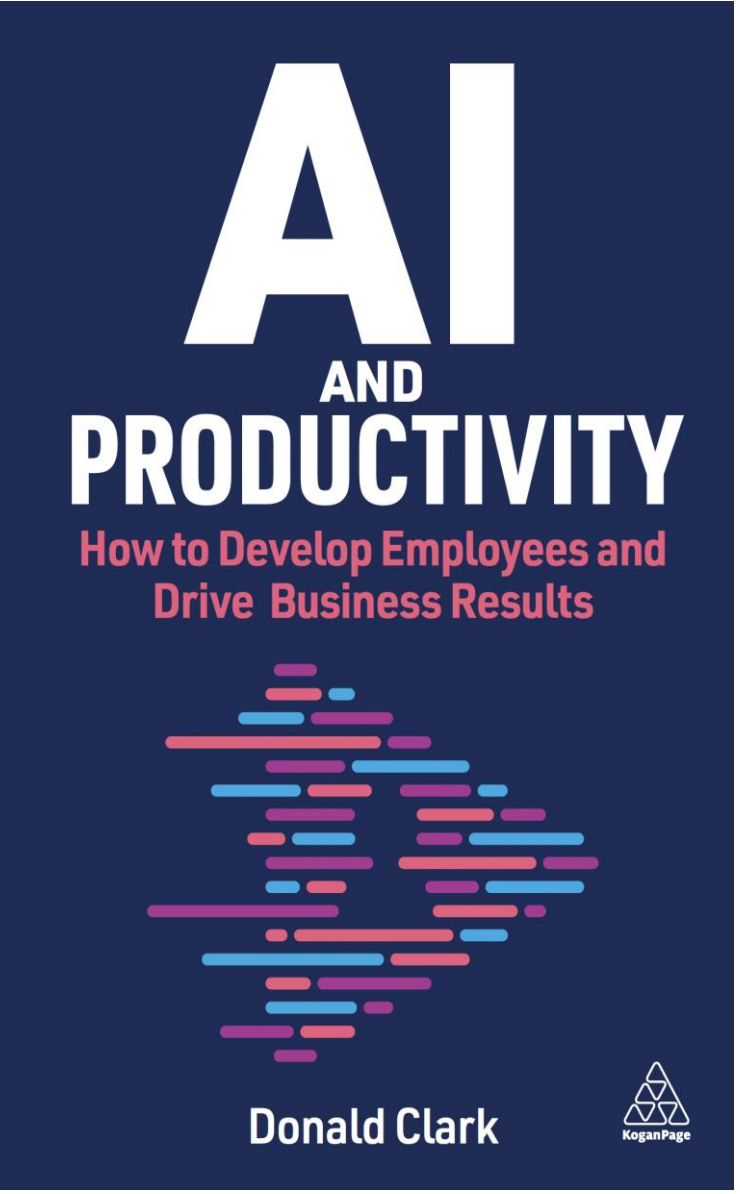
- AI Keynote or Conference
- AI Workshop/Hackathon
- AI Hub
- AI Champions & Ambassadors
- Brand your AI initiative



Deployment

- Direct access to a chatbot
- Pilots
- Wrappers with approved model(s)
- Custom AI assistants/agents
- Specific AI Agents
- AI-augmented workflows
- Early wins

Metrics



PHYSICAL productivity

Robots

Vehicles

Drones

Space

Vocational asssessment

Apprenticeship training & assessment

Entire world!

3 quotes ~**£500k**

Completed for £50k **(10% cost)**

6 months to **3 weeks**

36% increase in sales

62% could identify sale based on training

Pharmaceutical

Covid lab training

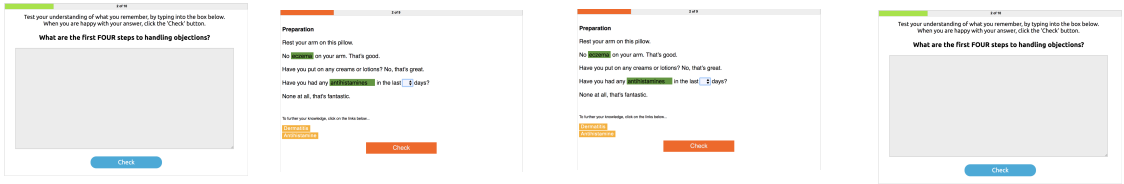
Had to be quick

Videos – pulsed

Open-input assessment

Results

2 weeks after video prod



Full license assessment

Covid assessment

1 of 7

2 of 7

Test your understanding of what you remember, by typing into the box below.
When you are happy with your answer, click the 'Check' button.

What do you wear when working within a safety cabinet?

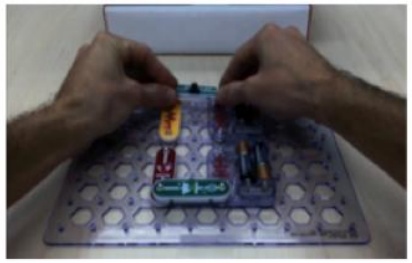
Check

Fast production

Low budget add on to video using AI

Which Perspective Leads to Better Learning of a Performance Task?

First-person perspective



Third-person perspective

Video producers:

- didn't like **chunking**
- had difficulty **editing for chunking**

Surgical support

https://www.youtube.com/watch?v=EE-TY-hwG4U&t=466s

Massive time saver on SMEs

SMEs don’t create but validate - loved it!

Quality the surprise result

Translated by AI to B1

Nurses loved it!

https://www.youtube.com/watch?v=EE-TY-hwG4U&t=466s

Training & Assessment platform

AI puts pedagogy into assessments

Generates TRAINING/**ASSESSMENTS** with **focus** on:

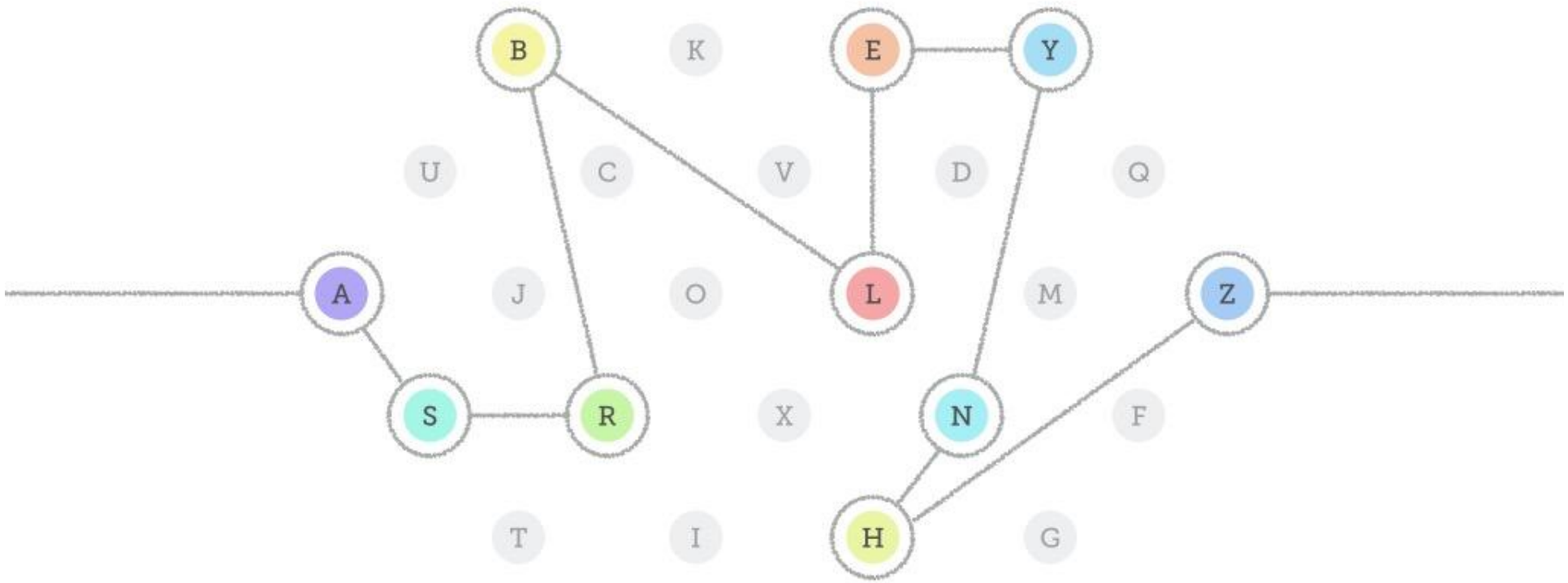
- ***Pedagogy of Transfer/Action***

....in any ***subject*** at any ***level***

In minutes not months

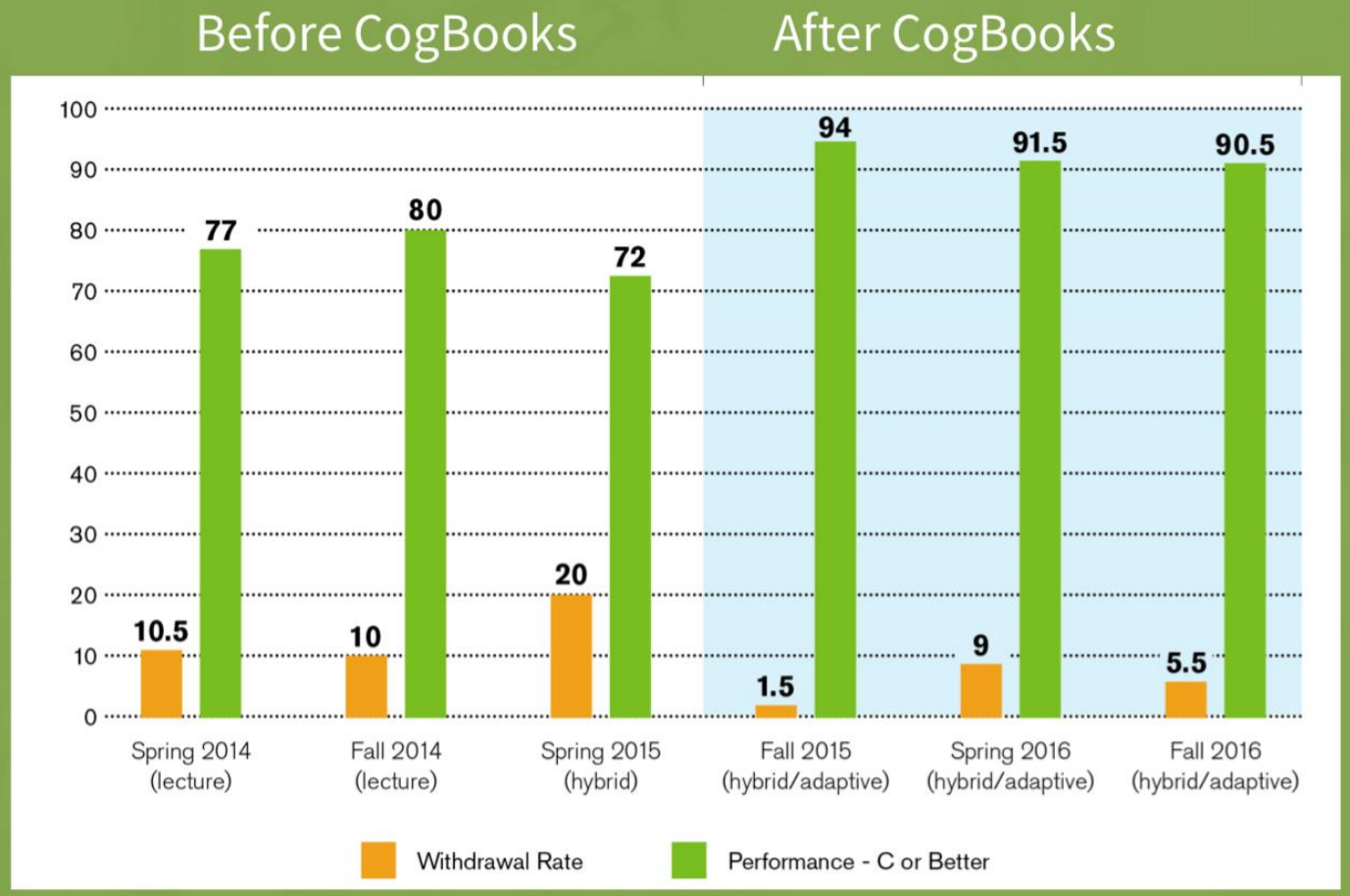
Adaptive learning company

Adaptive sets you free from the A to Z



It understands you – and finds the best way through

Courseware Success at ASU



20%
Improvement in pass rates

50%
Reduction in drop out rates

CogBooks™

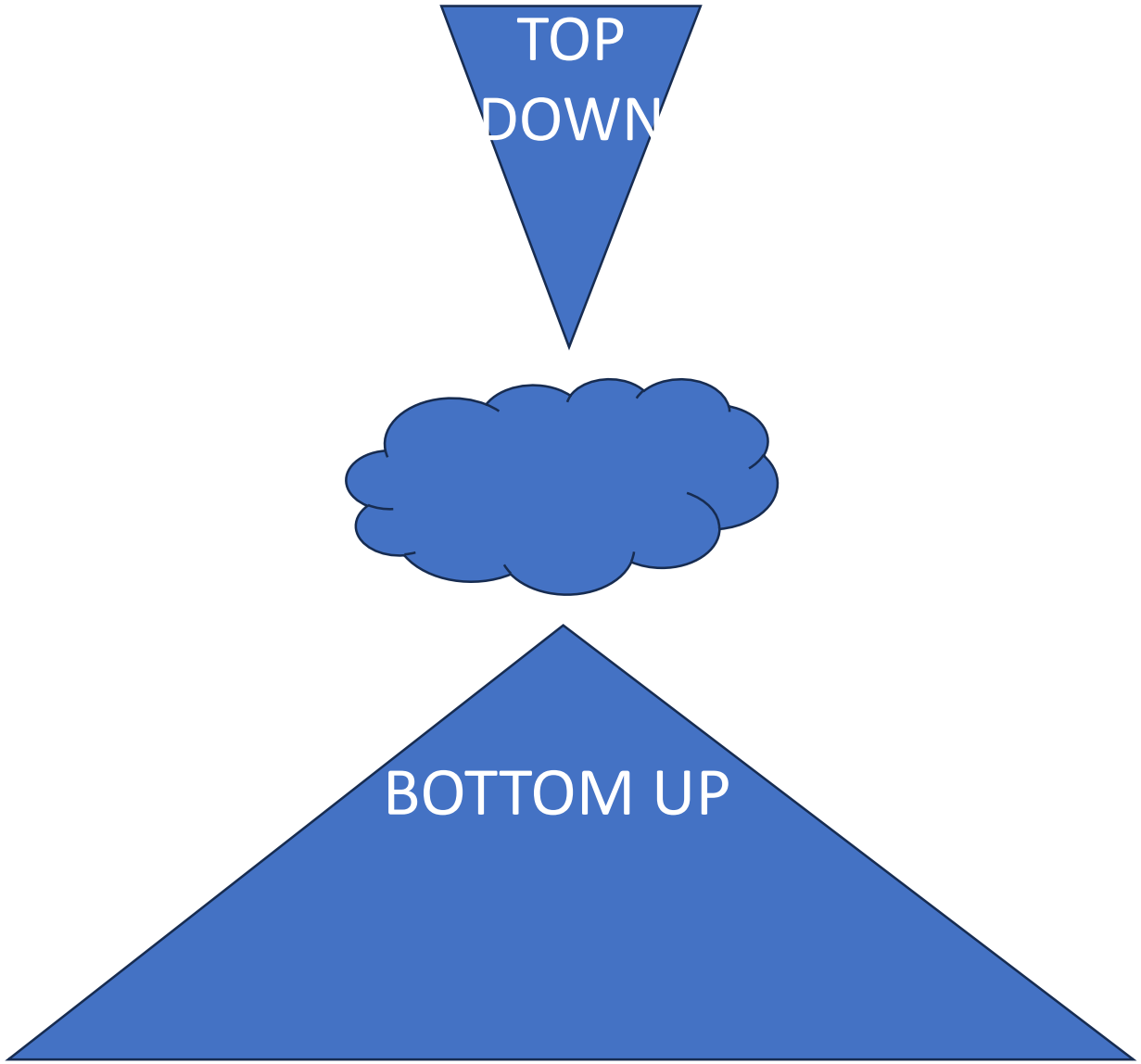
Great **learner evaluations**
Improved learning outcomes

Deep-seated **faculty resistance**
Research doesn't matter!
Funding ends, project ends
Mosquito v tortoise projects

Global Manufacturing

AI first approach

CEO - driven



AI first approach

Top-down

CEO - driven

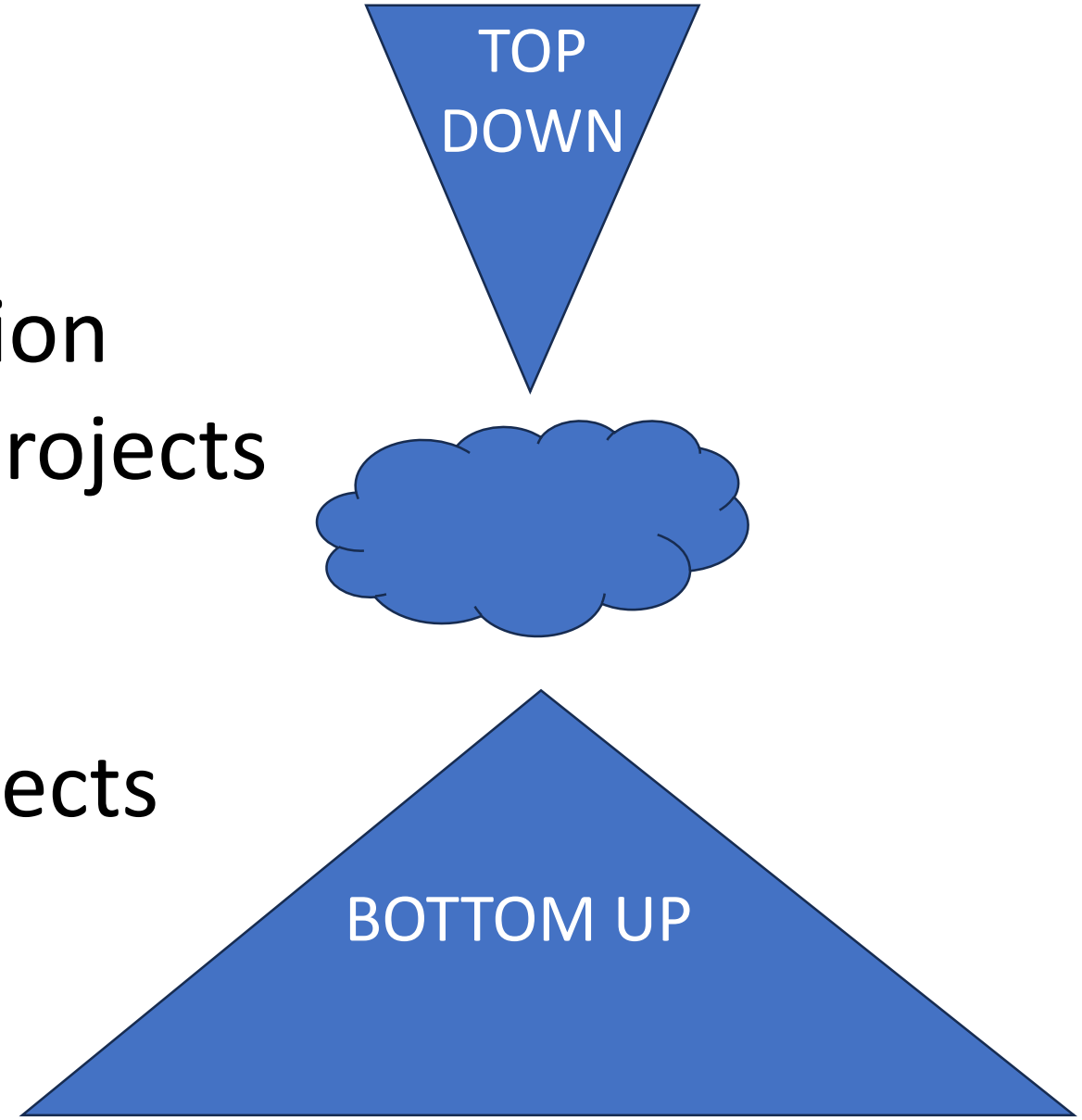
Initial '**urgency**' presentation

Senior Leadership Team projects

Funded real projects

Students to complete projects

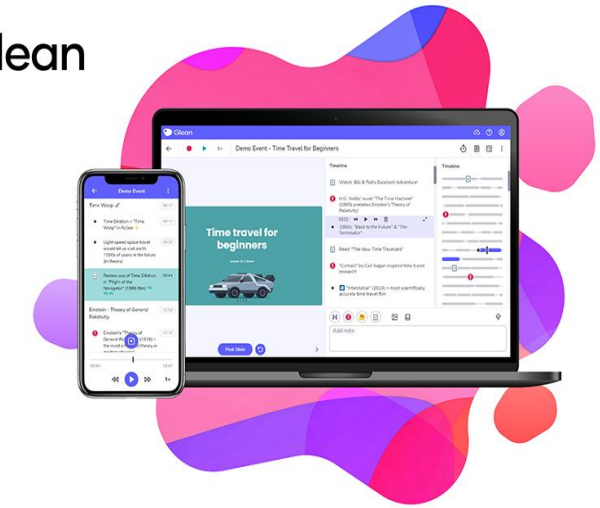
No metrics for success



Higher Education

Feedback on note taking with AI

Glean



Compares **notes** with **lecturer transcript**

Automatically splits transcripts & adds relevant headings

Clustered notes – notes to right place

QuizMe - automatically **adds assessment**

Great **learner feedback**

Great **customer feedback**

Some fierce resistance from HE

- Assumed **data used for training** by OpenAI
- Assumed company **training own model**
- **Abstract philosophical debate on ethics**

Difficult to get access to user **data**

Artificial cost constraints on **scaling**

Training platform

Action/transfer pedagogy built-in
Full product built
Award winning

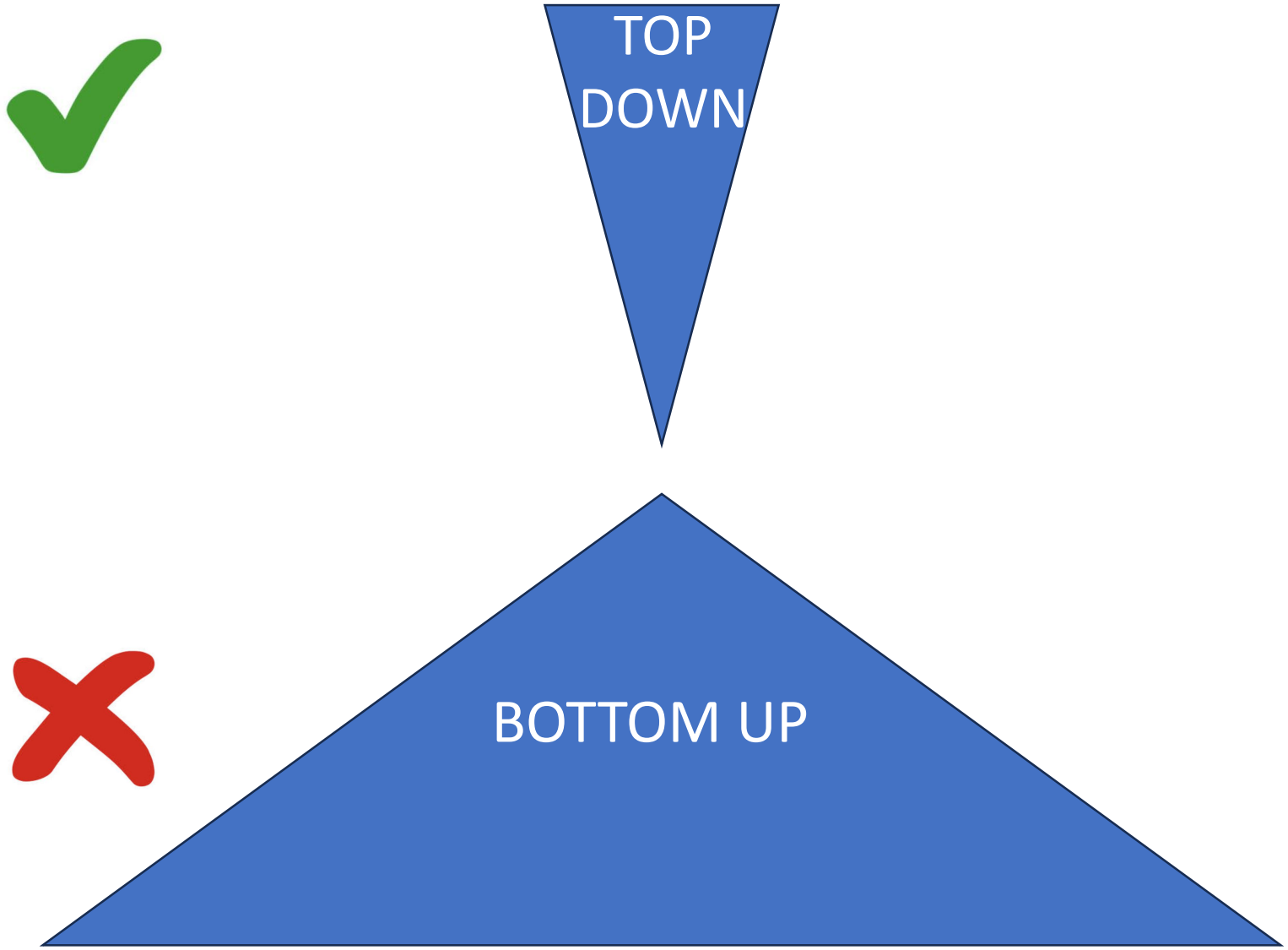
Low knowledge of AI in L&D team
Traditional L&D wanted traditional features
"It would nice to have this…"
Drift backwards
EU forced academic input costly & no impact
Unaligned marketing – **no product name**

Publisher

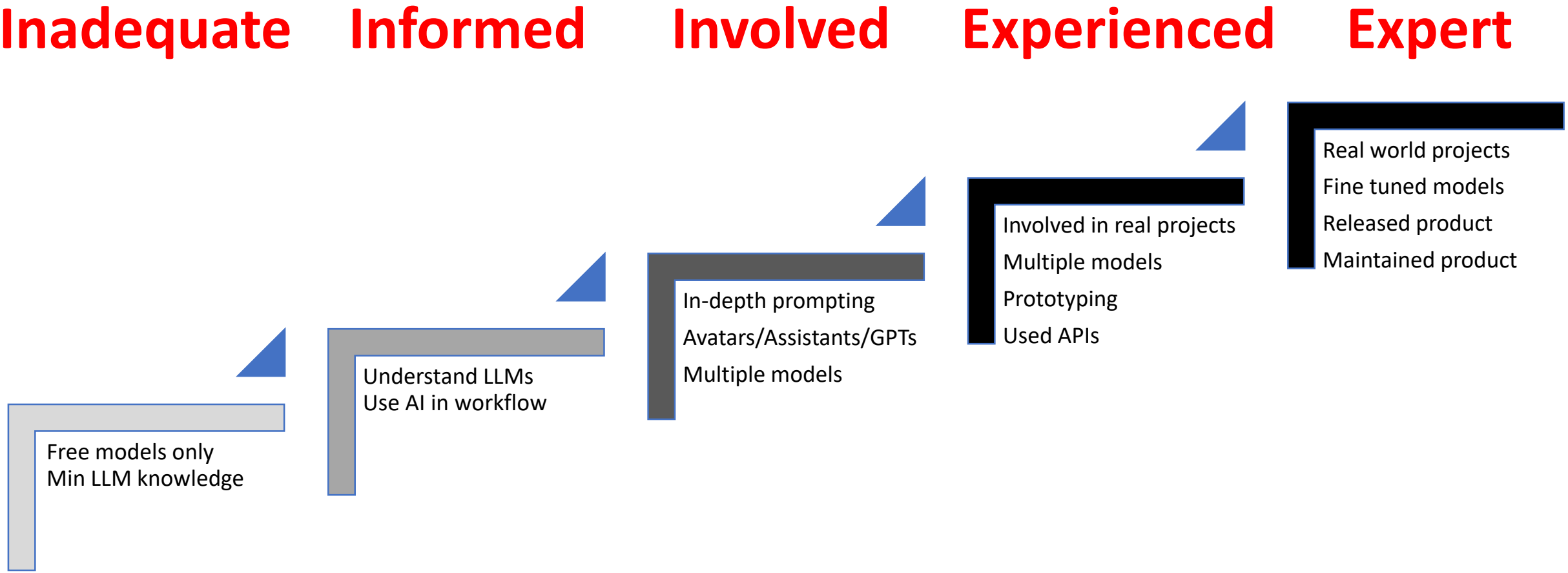
Language level worked
Edit source material (own documentation)

Expectations too high (AGI chatbot)
Publishing process not appropriate (AI disruptive)
Old e-learning mindset & methodology
Low knowledge of AI a problem
Team too big (marketing etc.)
Butterfly effect… saw this on internet…

Implementation



Levels of LLM expertise (high as possible)



Top down support

Give project a **name**

Define of **goals** (not chatbot)

Small team

AI mindset in full team:

- Expertise matters
- Understand probabilistic AI
- Fluid & fast moving tech
- Data issues

Built-in **pedagogy**

Manage **expectations**

Move **fast & prototype** (a lot)

Edit **source material**

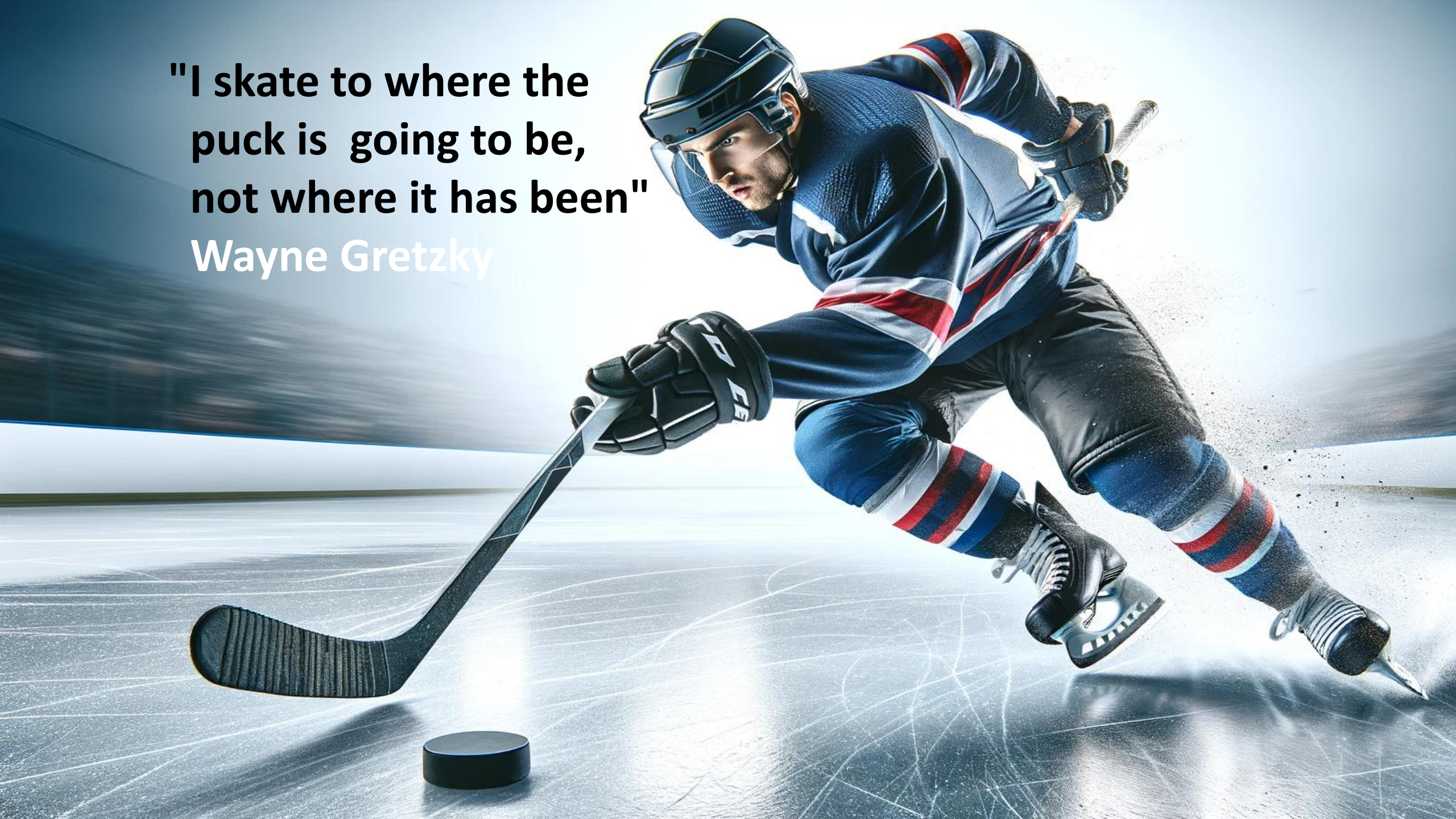
Own **documentation**

SMEs **validate**

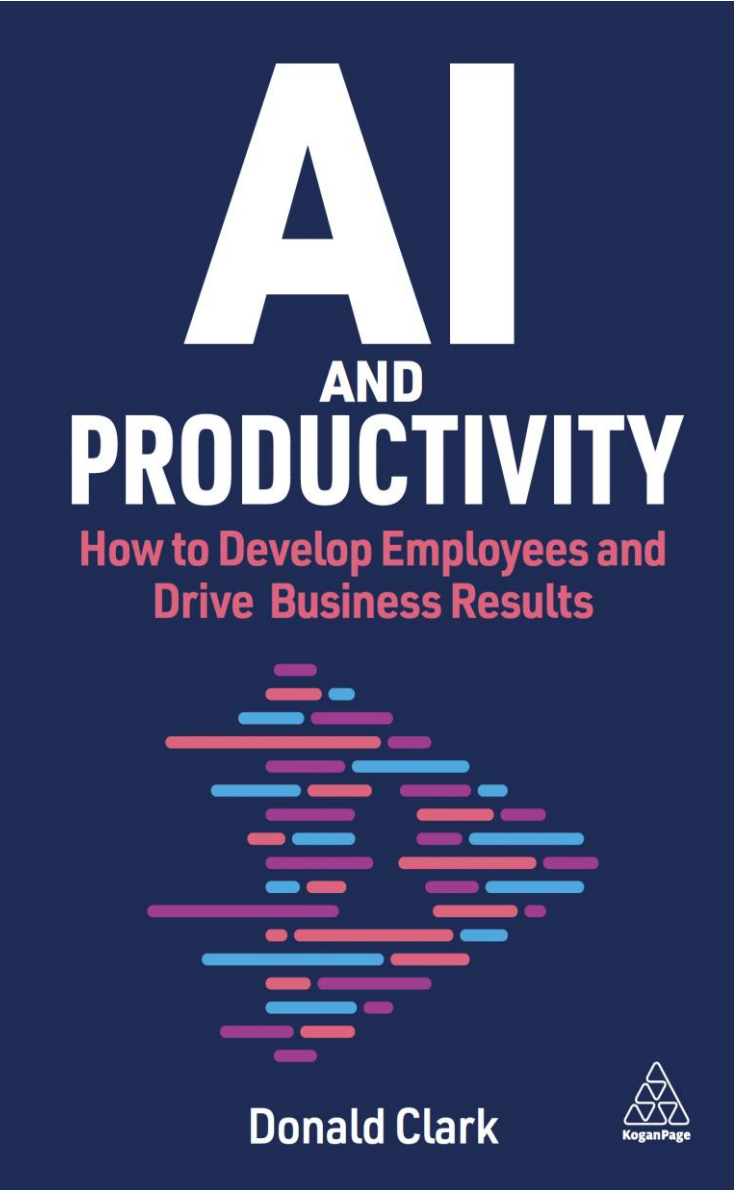
Employ **risk management**

Legal/data constraints

Push through prototype to **produc**


"I skate to where the
puck is going to be,
not where it has been"
Wayne Gretzky

Thanks for listening



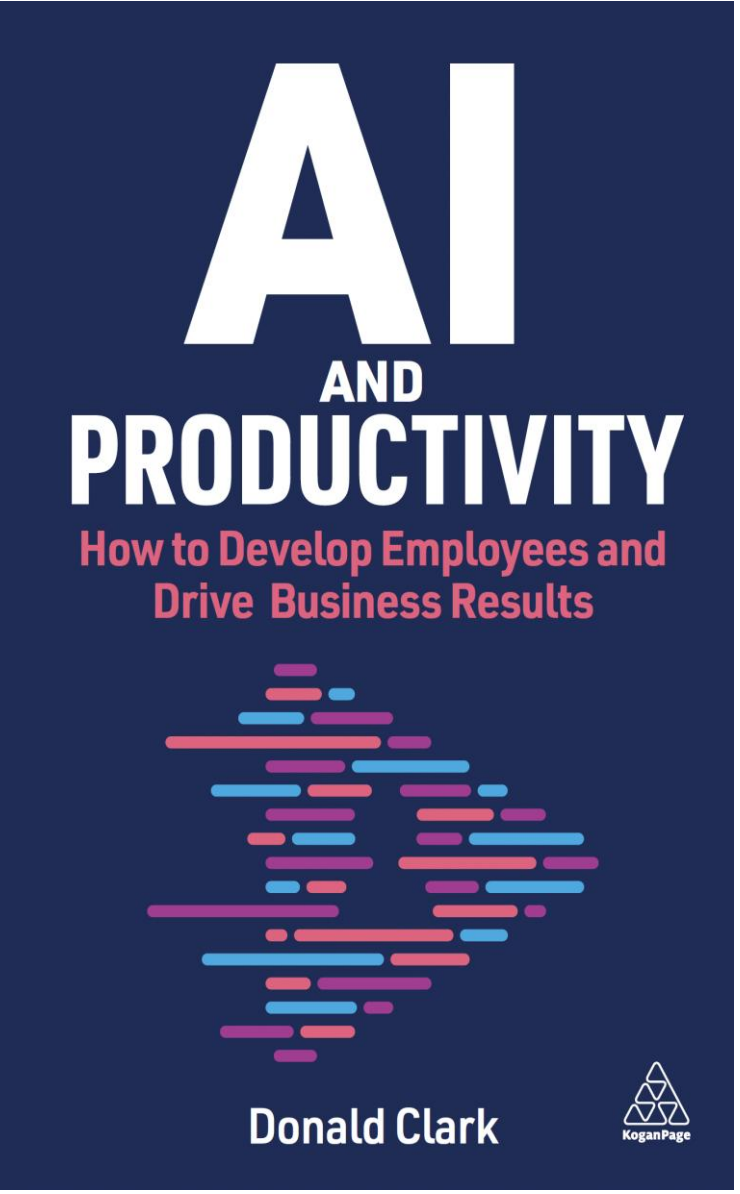
Part 1
- What is PRODUCTIVITY?
- EVIDENCE for productivity
- BRAINS, AGENTS & AGI

Part 2
- Productivity PARADOXES
- HUMANS-IN-THE-LOOP
- PARTNERS in productivity

Part 3
- MULTIMODALITY
- HEALTHCARE productivity
- PHYSICAL productivity

Part 4
- GETTING STARTED
- DEPLOYMENT IN ORGs.
- MEASURING productivity

Part 5
- ETHICS
- FUTURE



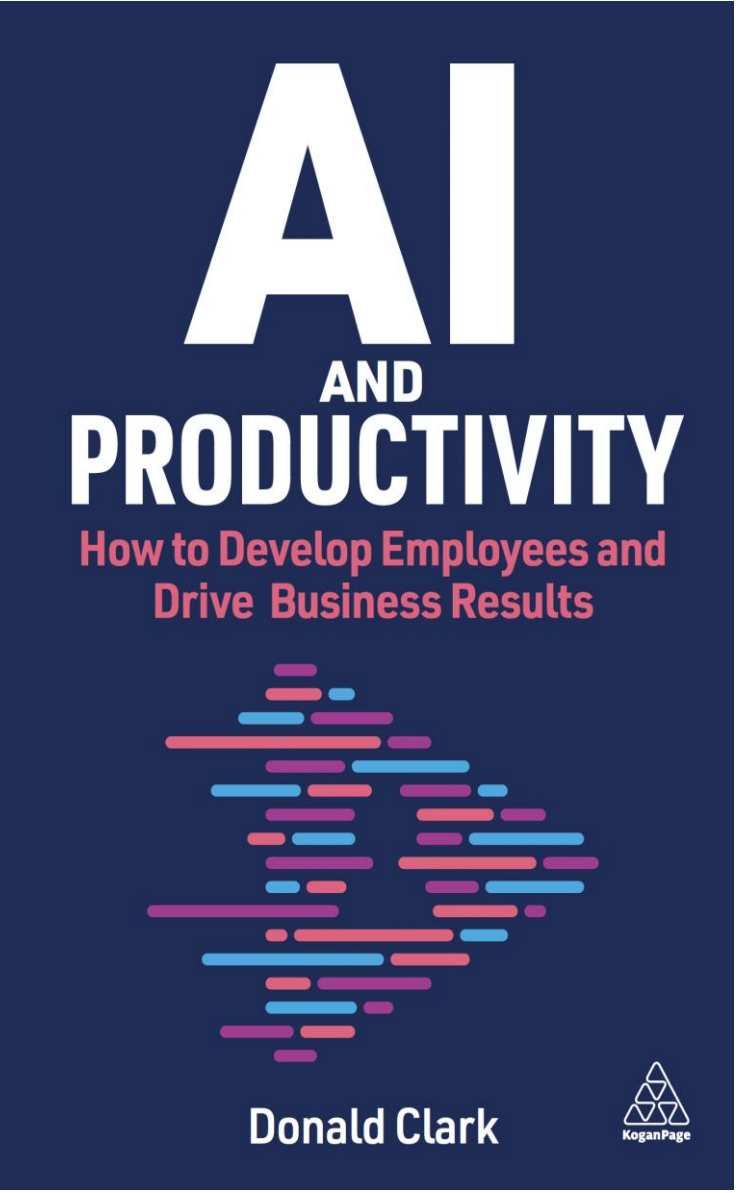
What is **PRODUCTIVITY**?

"*Productivity isn't everything, but in the long run, it is almost everything. A country's ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker.*"

Paul Krugman

Technological escapism

Economic Productivity	National	**Productivity Plus**	Innovative
	Sectoral		Physical
	Organisational		Cultural
	Team		Ecological
	Personal		
	Task		



What is **PRODUCTIVITY**? GDP, Sector, Org, Team, Personal & MORE

EVIDENCE Academic, field research, reports, surveys

BRAINS, AGENTS & AGI Cog capped, Agent Types, Issues, AGI, ASI

Productivity **PARADOXES** People, Org, Tech, Economic

HUMANS-IN-THE-LOOP In/On/Out/Experts/AI/Centaurs/Cyborgs

PARTNERS in productivity Creativity, Critical, Comms, Collaboration

MULTIMODALITY. Text, Images, Audio, Video, Art

HEALTHCARE Triage, Clinical decision making, Research

PHYSICAL productivity Robots, Vehicles, Drones, Space

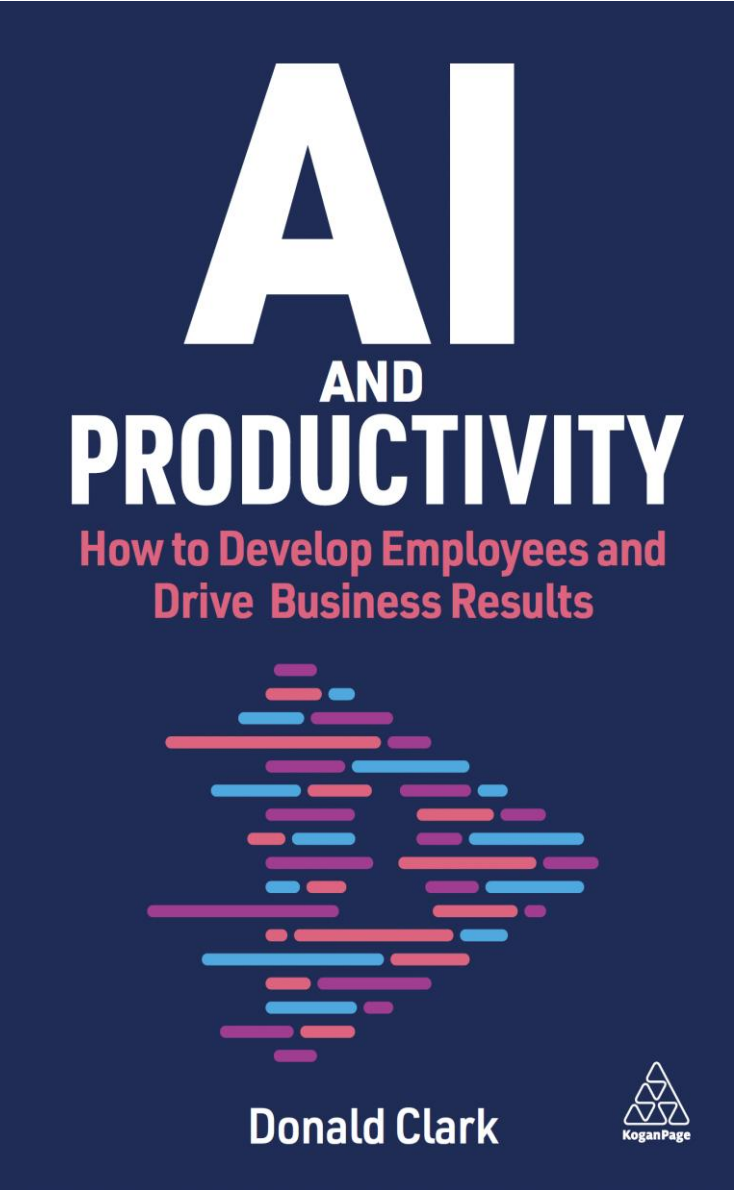
GETTING STARTED Policies, Prompting, Agents/workflow

DEPLOYMENT IN ORGs. Top/bottom, change, types, methods

MEASURING productivity Formula

ETHICS Employment, Meaning

FUTURE Research, Runaway, Holy trinity!



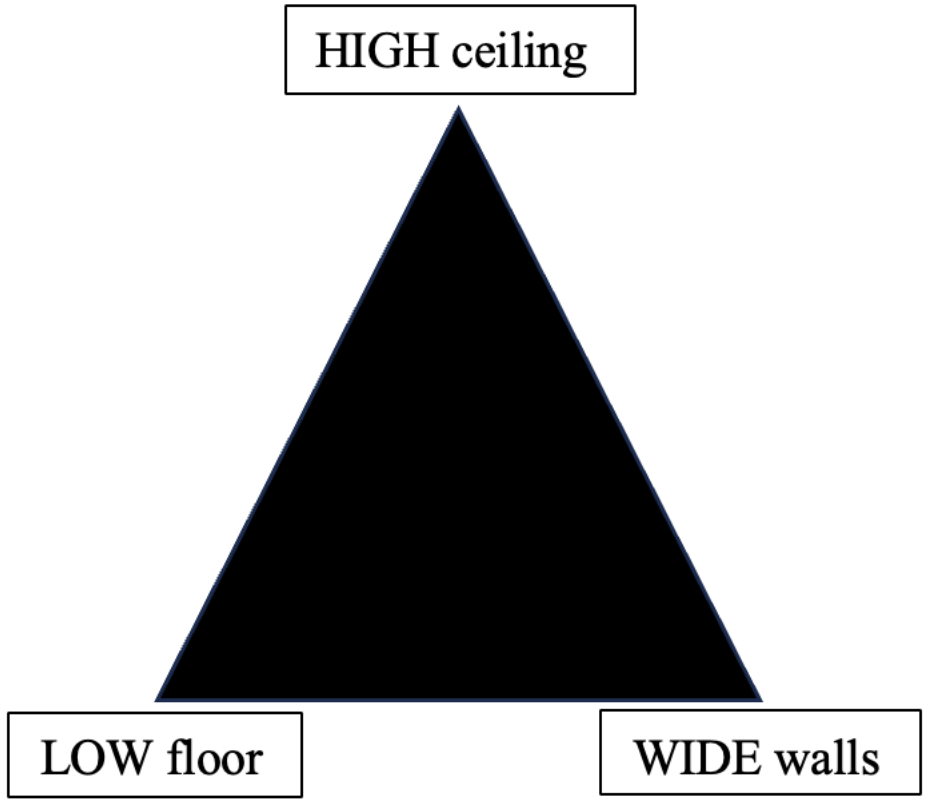
PARTNERS in productivity

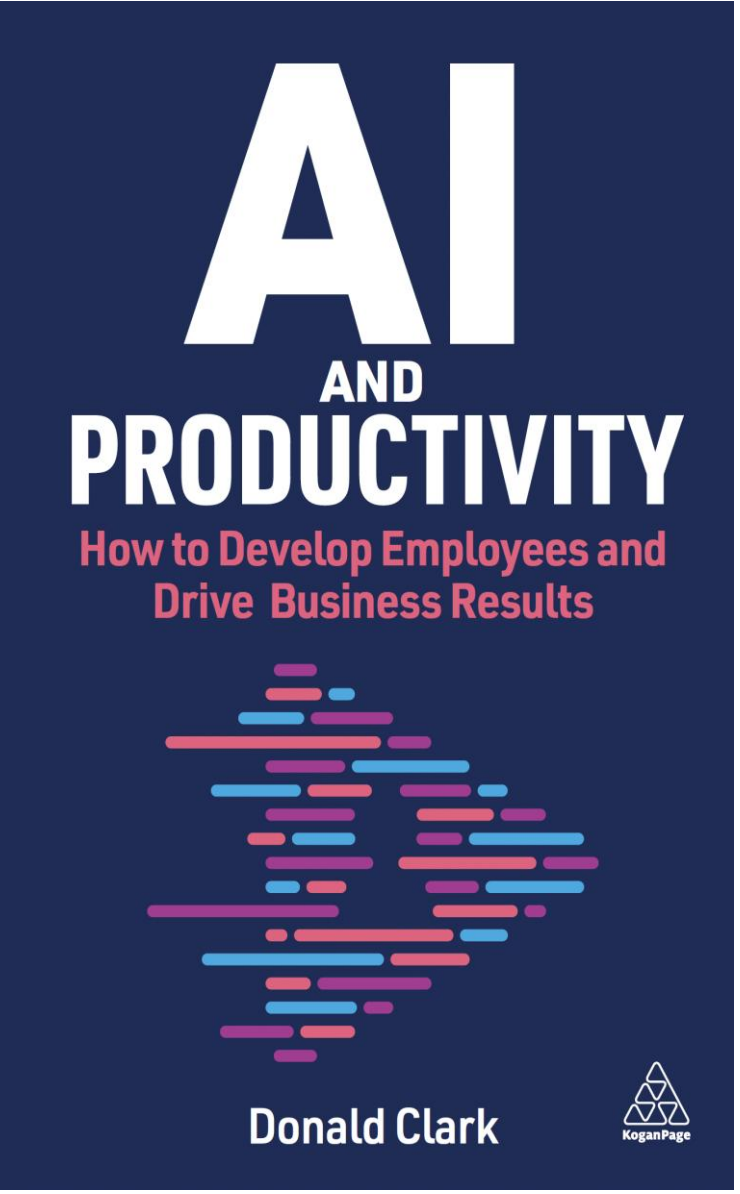
False dichotomy

Paradox of human exceptionalism

- Creativity
- Critical thinking
- Communication
- Collaboration

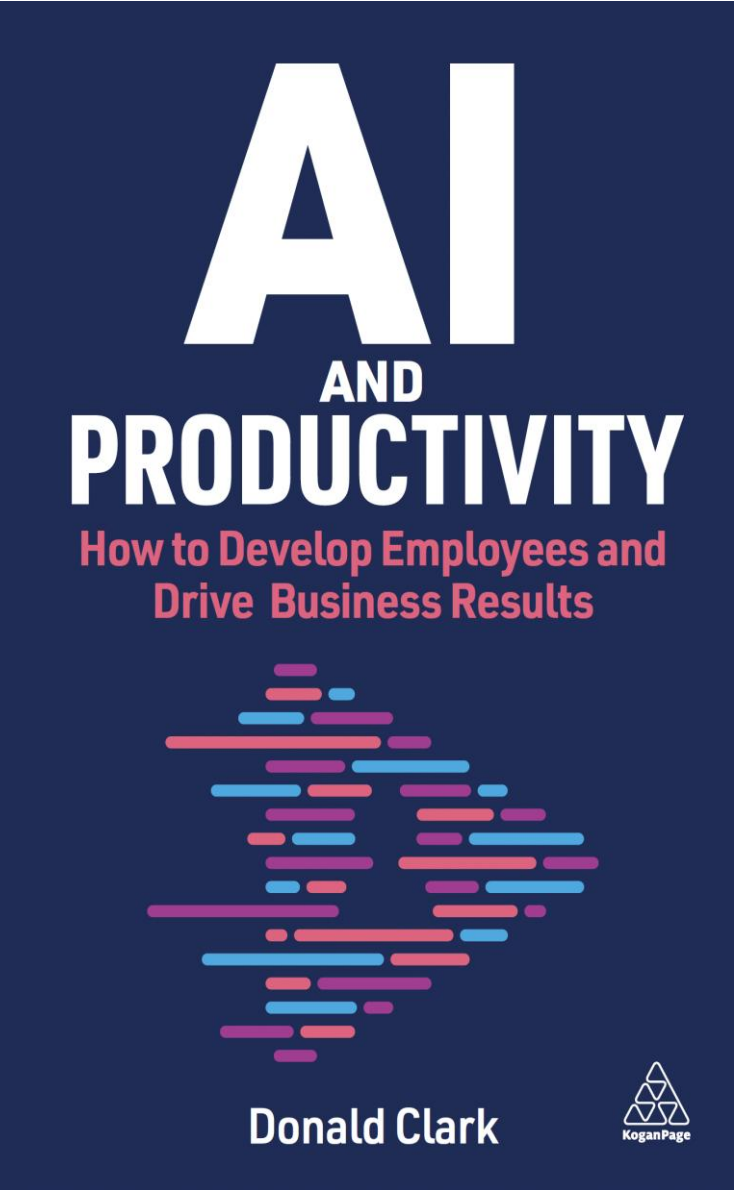
New paradigm





HEALTHCARE

Prevention
Presentation
Diagnosis
Investigation
Treatment
Administration
Wellbeing
Training



GETTING STARTED

Policies, Prompting, Agents/workflow

Flipped the script (Andrej Karpathy)

Policies and inaction

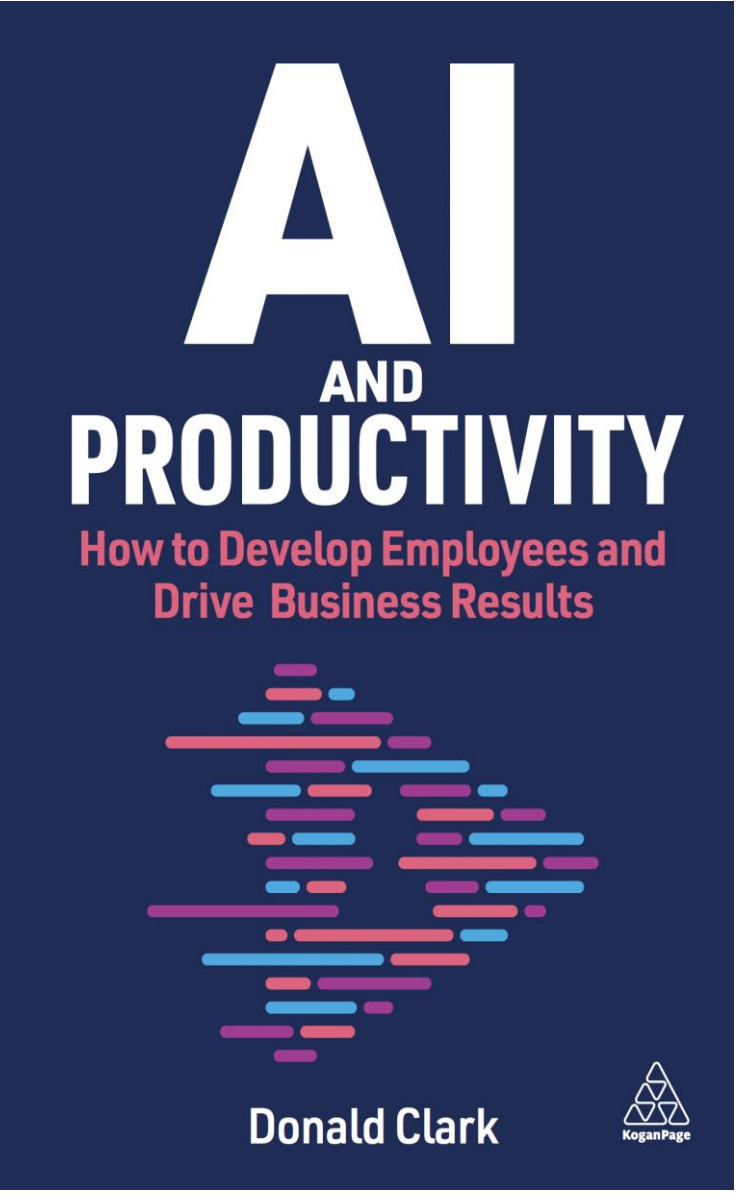
Productive Prompts

- Prompt libraries & AI templates
- Unified prompt framework
- Advice

Experts-in-the-loop

Prompting agents

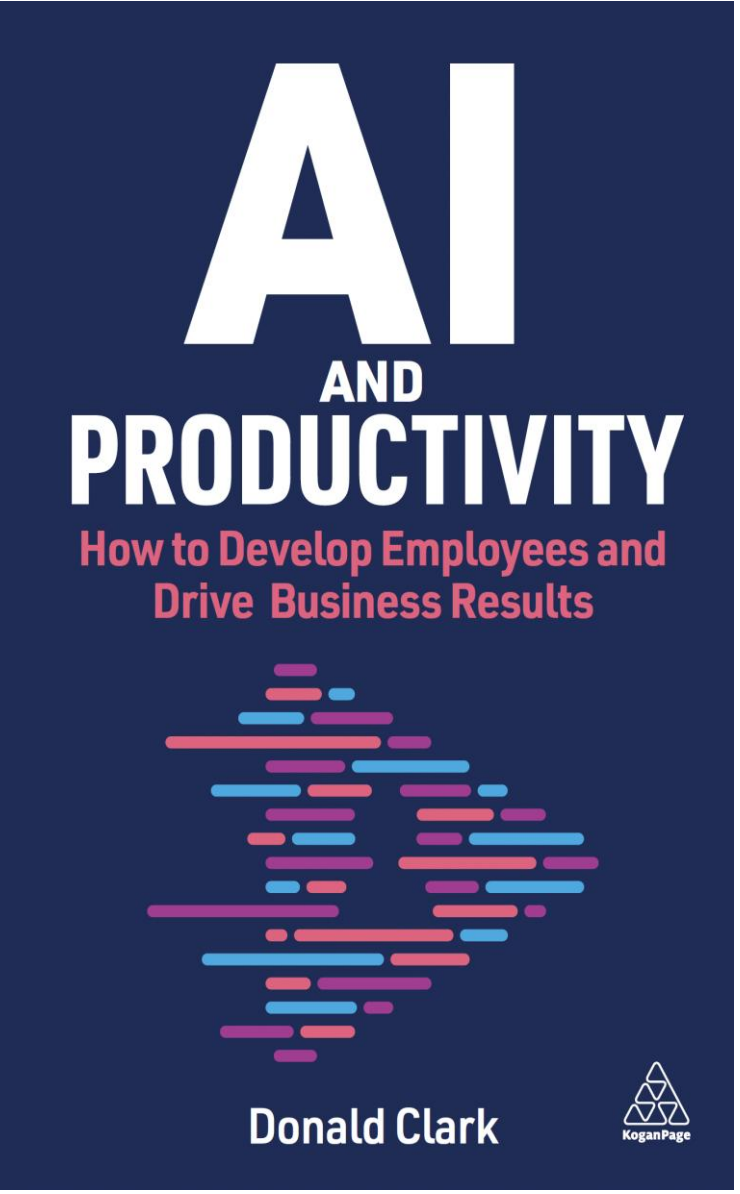
Policies prevent action!



MEASURING productivity

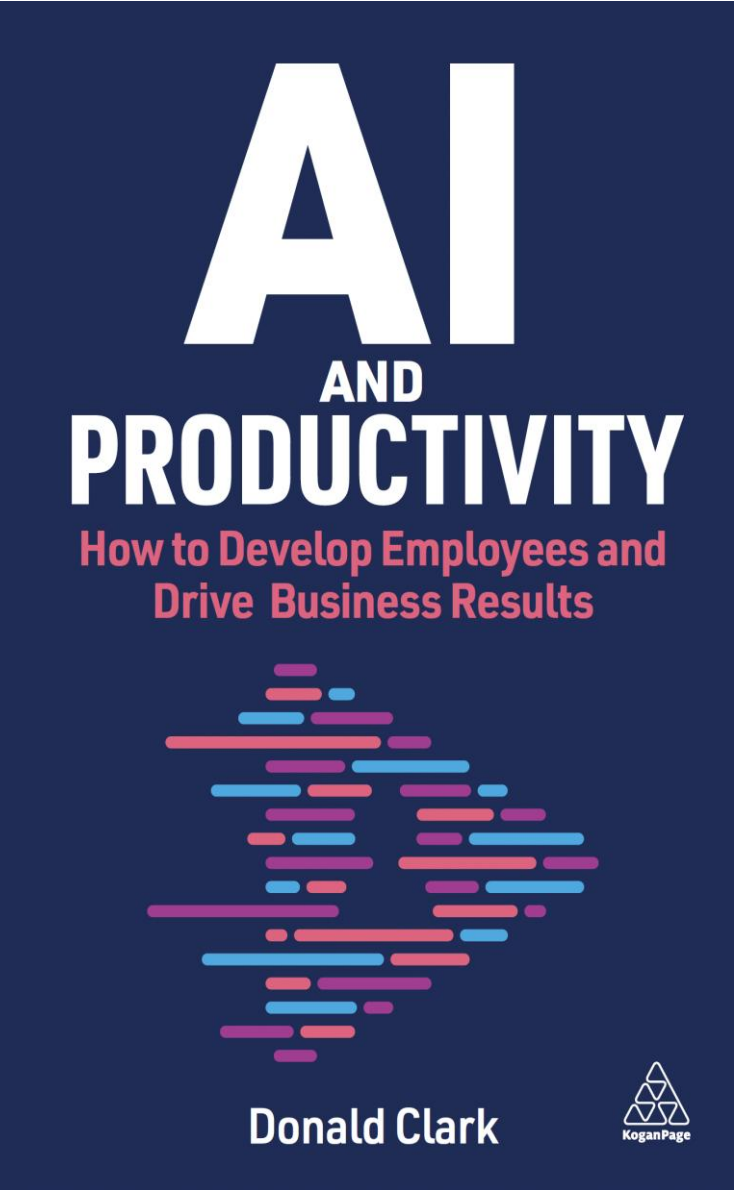
Metrics and formula
Using AI data analytics

We use the wrong metrics!



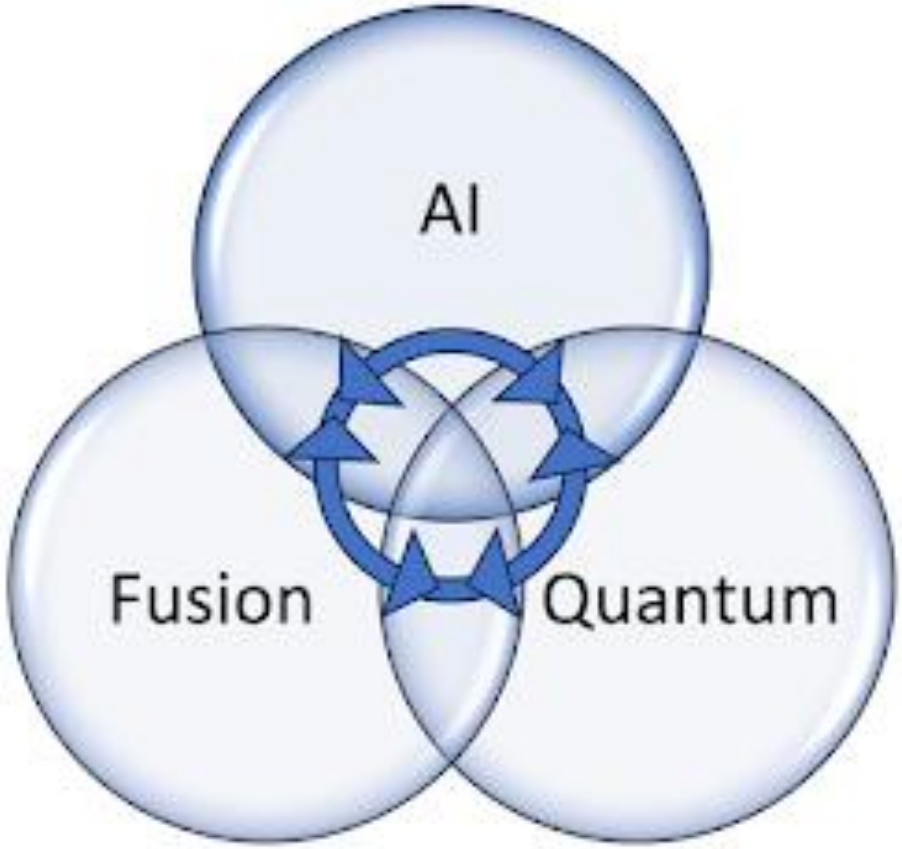
ETHICS of productivity

Plagiarism	Cheating and copyright
Prejudice	Biases
Persuasion	Being duped
Privacy	Data and security
Passivity	Dehumanisation and skills atrophy
Partisan	Digital divide
Poverty	Immiseration and unemployment
Planet	Energy and emissions
Perish	Extinction event



FUTURE

Radical Productivity
Autogenesis
Science
Biology
Chemistry
Physics



Mosquito v tortoise projects

Funding ends, **project ends**

Beware of **rising tide** projects

Expectations too high

All-encompassing chatbot

Low knowledge of AI a problem

Traditional **L&D/e-learning** mindsets

Using **old project/design** processes

Deep-seated **faculty resistance**

Educators want **control** but need to let go

EU academic input costly & no impact

Ask **customers** for design ideas (feedback)

Ask **L&D/faculty** for features

Unaligned marketing – **no product name**

Scope creep

Underestimated **testing**

Difficult to get access to **data**

Ethical policies/frameworks way too abstract

Data myths

- Assumed **data used for training** by OpenAI
- Assumed company **training own model**

Abstract policies/debate on ethics

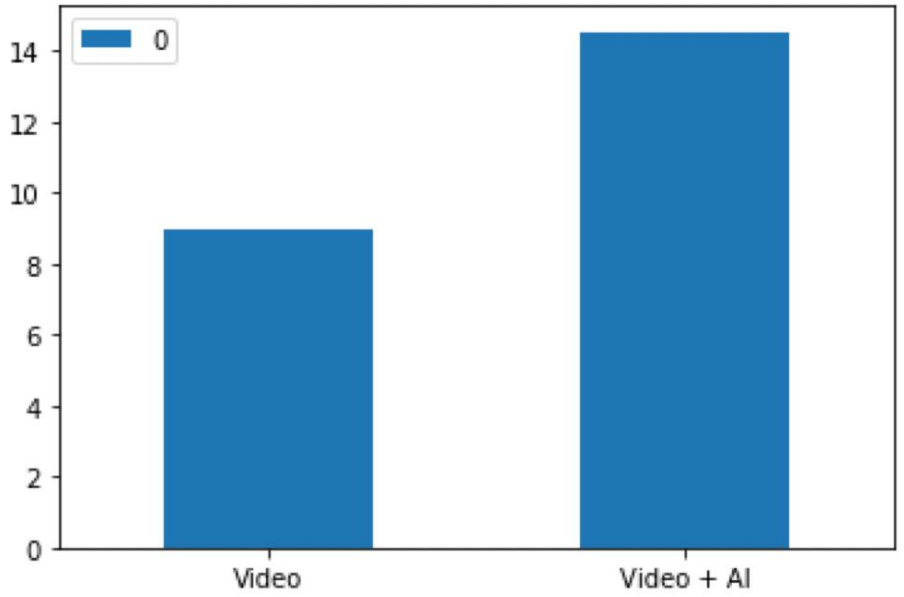
Confusion on cost constraints on **scaling**

Butterfly – saw on internet last night…

Don’t fine tune your **own models**

VIDEO v VIDEO + WildFire

Figure 1: Histogram comparing mean scores of each group

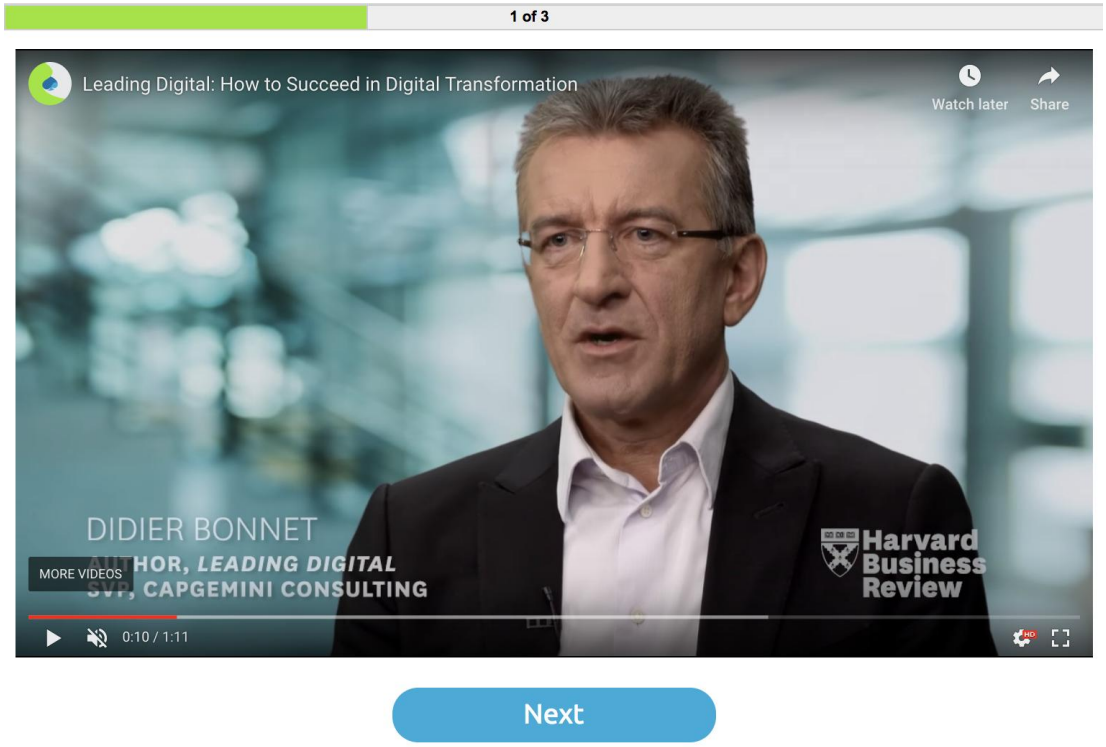


HUGE increases in retention

61.5% increase in mean retention, from a mean value of 9.00 to 14.54

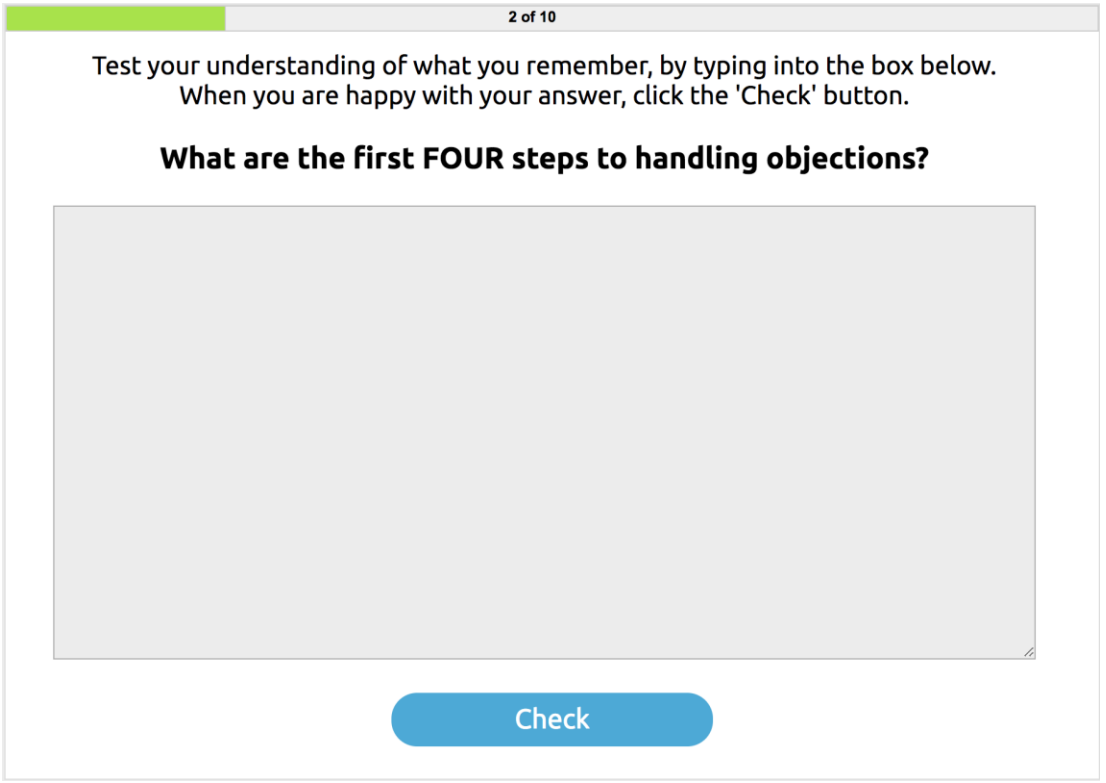
Consultancy

Conference spend ineffective!



Online assessment

Automated question creation



Open input

With marking & feedback

Voice

Fast production

Low budget add on to video

Higher retention

Consultants think they already know

Learning not seen a serious activity

Telecommunications

Recruitment training

Expert system

Knowledge of employment law

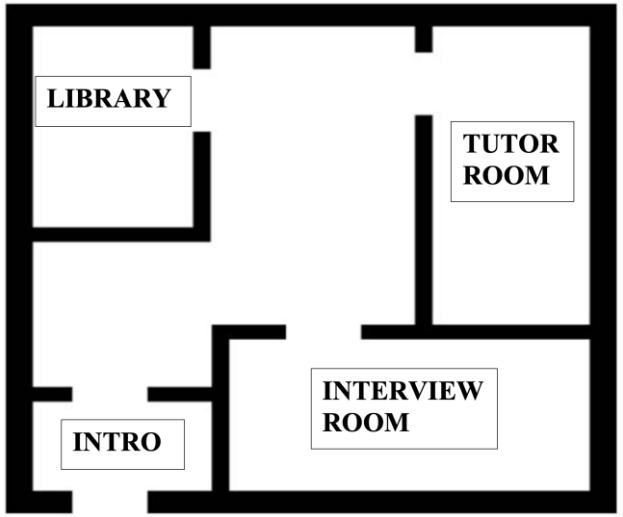
Score/select CVs

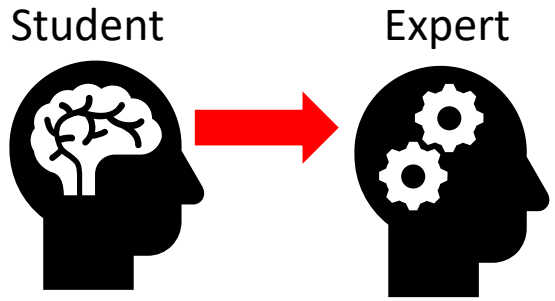
8 full interview simulations (random)

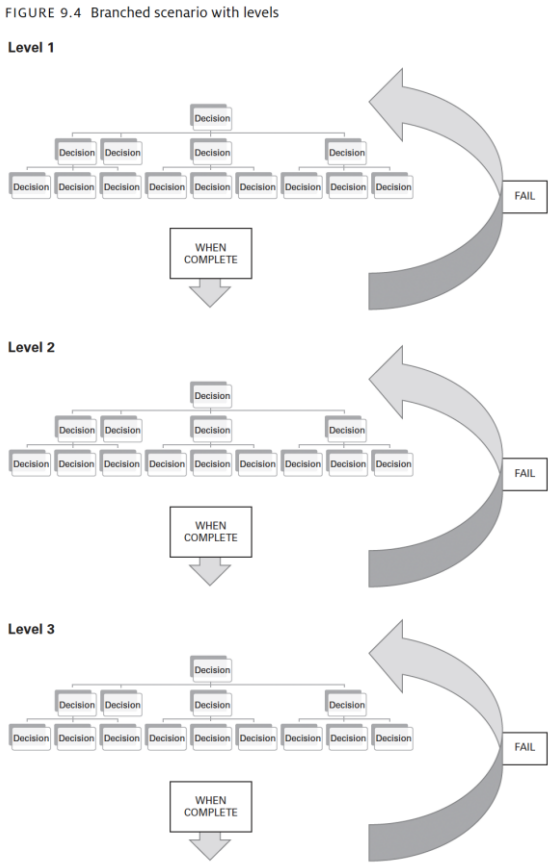
Results

Suspension of disbelief

High skills acquisition & assessment







NHS

1 of 9

Next

2 of 9

Preparation

Rest your arm on this pillow.

No eczema on your arm. That's good.

Have you put on any creams or lotions? No, that's great.

Have you had any antihistamines in the last days?

None at all, that's fantastic.

To further your knowledge, click on the links below...

Dermatitis
Antihistamine

Check

2 of 10

Test your understanding of what you remember, by typing into the box below.
When you are happy with your answer, click the 'Check' button.

What are the first FOUR steps to handling objections?

Check

Fast production
Low budget add on to video

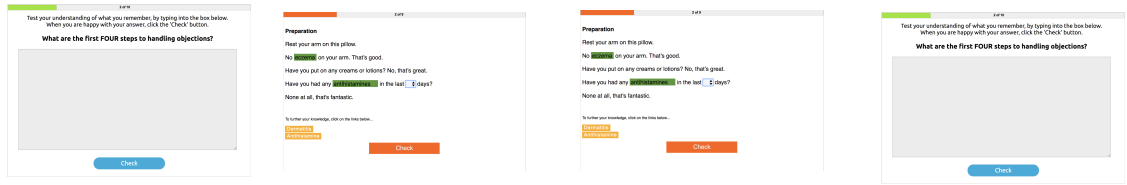
AI makes errors BLOW!
Tolerance for failure

Construction

2018

Protected characteristics

- Age
- Disability
- Gender reassignment
- Marriage or civil partnership
- Pregnancy and maternity
- Race
- Religion or belief
- Sex
- Sexual orientation



Recruitment training

Knowledge of employment law

Score/select CVs

8 full interview simulations (random)

Results

Suspension of disbelief

High skills acquisition & assessment

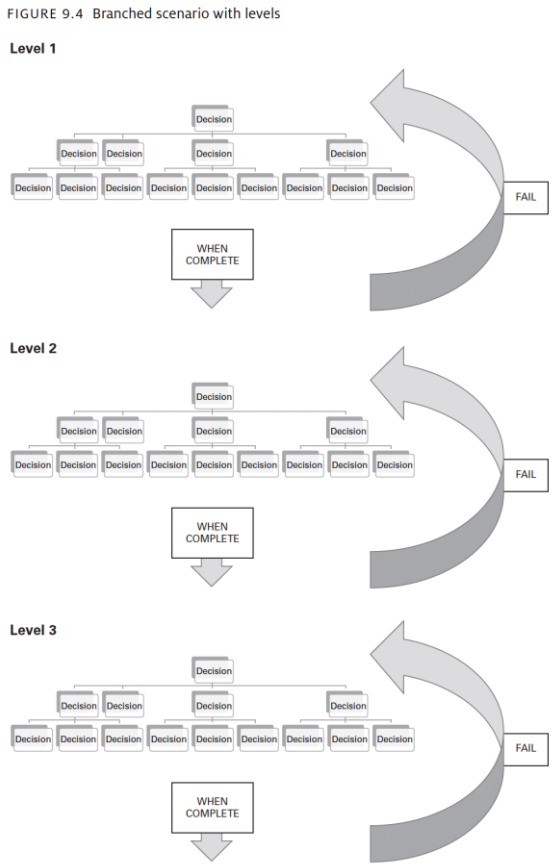


FIGURE 9.4 Branched scenario with levels

Marking short answers with AI

Table 3: Model Evaluation of Student Answers

	GPT 3.5 Zero-Shot	GPT 3.5 Few-Shot	GPT 4 Zero-Shot	GPT 4 Few-Shot
Correctly Predicted	71 %	72 %	85 %	84%
Precision	0.94	0.93	0.85	0.87
Recall	0.50	0.51	0.85	0.85
F1	0.65	0.66	0.85	0.86
Percent Agreement	0.71	0.72	0.84	0.85
Cohen's Kappa	0.44	0.45	0.68	0.70

STUDY

AI to assess **open text answers**

Series of experiments

Different domains (Science & History)

Grade levels **5-16 (Key Stage 2/3/4)**

RESULTS

GPT-4 performed well (0.70)

Similar to human performance (0.75)

Consistent with previous research

AI **as good as humans**

Henkel, O et al. July 2024, Can Large Language Models Make the Grade? An Empirical Study Evaluating LLMs Ability To Mark Short Answer Questions in K-12 Education. In *Proceedings of the Eleventh ACM Conference on Learning*.

2025 Year of agents, multimodality & reasoning

Intelligent agents

- **Create assessments**
- **Assessors**
- Real-world **scenarios**
- **Authentic** assessments

Multimodality

- Written, oral & video
- **Reflective** assessment
- Richer evaluation of **competencies**
- **Holistic, richer, authentic**

Advanced reasoning

- Sophisticated **feedback**
- **Critiques**
- **Critical thinking**

Rising tide!