

Strategic AI opportunities for non-profits and cake

27 February 2025



About Adapta



- We are a **specialist** information systems consultancy
- We only work with **charities**, associations, arts and culture organisations and others in the not-for-profit sector
- We are completely **supplier-independent**
- Our consultants have held **senior** positions in a broad range of different organisations
- Our advice and guidance is based on **practical experience** gained over many years

About Adapta



Your Objectives and Plans

Strategy and roadmap

Reviews

Leadership and advice

Governance and structure

Your Organisational Capabilities

Technology solutions and partners

Risk and compliance

Workforce

Your Success

Implementation and change delivery

Technology adoption

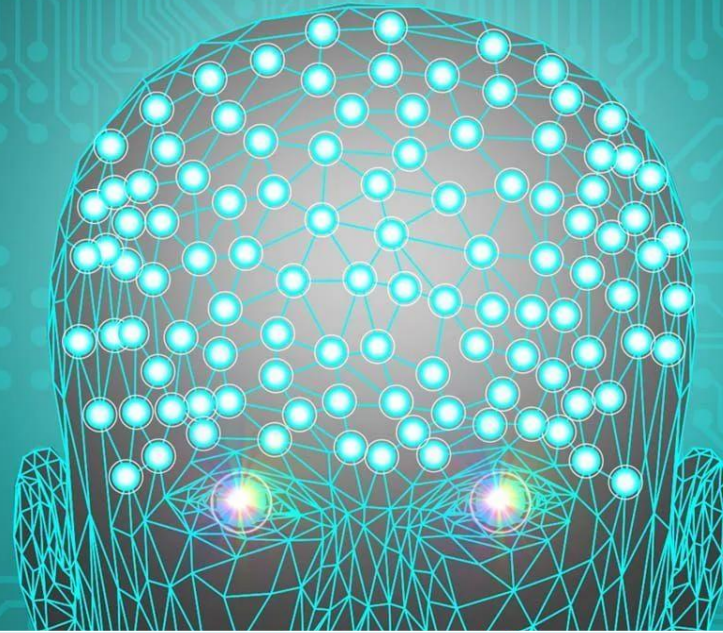
Programme



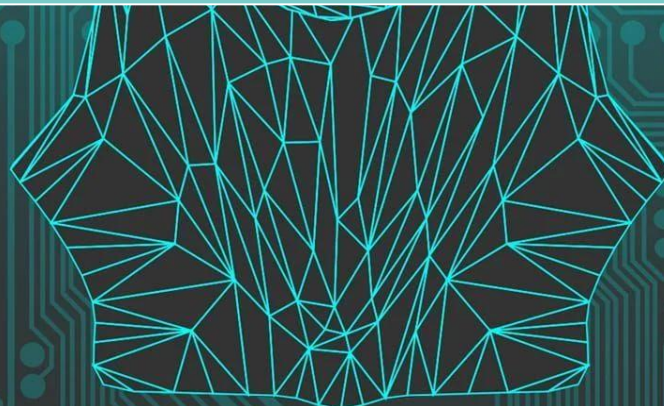
- 14:00 **Arrival and welcome**
Welcome - introductions and overview of the agenda for the afternoon.
- 14:10 **Presentation 1 – AI can do anything right?**
A helicopter overview of how AI is or can be practically used in the third sector today.
- 15.00 **Table session 1 – Strategic opportunities for AI**
Table discussions on attendee’s thinking on the most valuable ways AI could help their organisation
In the next 3 years.
- 15.30 **Tea/Coffee – opportunity to network**
- 15.45 **Presentation 2 – The AI marketplace for third sector organisations**
A tour of the AI marketplace, what market options might be relevant for you and what organisational capabilities you might need to adopt.
- 16.15 **Table session 2 – Adopting AI**
Table discussions on what solutions have been employed to date and what has been learnt in respect of organisational readiness and adoption.
- 16-45-
17.00 **Review and close**



Introductions



**What do you mean
“Strategic AI Opportunities?”?**



Perspectives on Human Intelligence



1 Gardner's 9 Intelligence Types

Gardner's theory identifies nine distinct intelligence types, including linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, naturalist, and existential.

2 Emotional Intelligence (EQ)

EQ encompasses skills like self-awareness, empathy, and social skills.

3 Cultural Intelligence (CQ)

CQ is about awareness of and navigating cultural differences.

4 Artificial Intelligence (AGI)

AGI in theory encompasses all the above; in practice AI today – and strategically over the next 3 years – is more narrow in focus.

Neuro and Cognitive Diversity

Impact on Team Performance

Research shows that such diversity can enhance team performance at all levels of an organization from trustees down.

AI offers cognitive capabilities your staff may not have

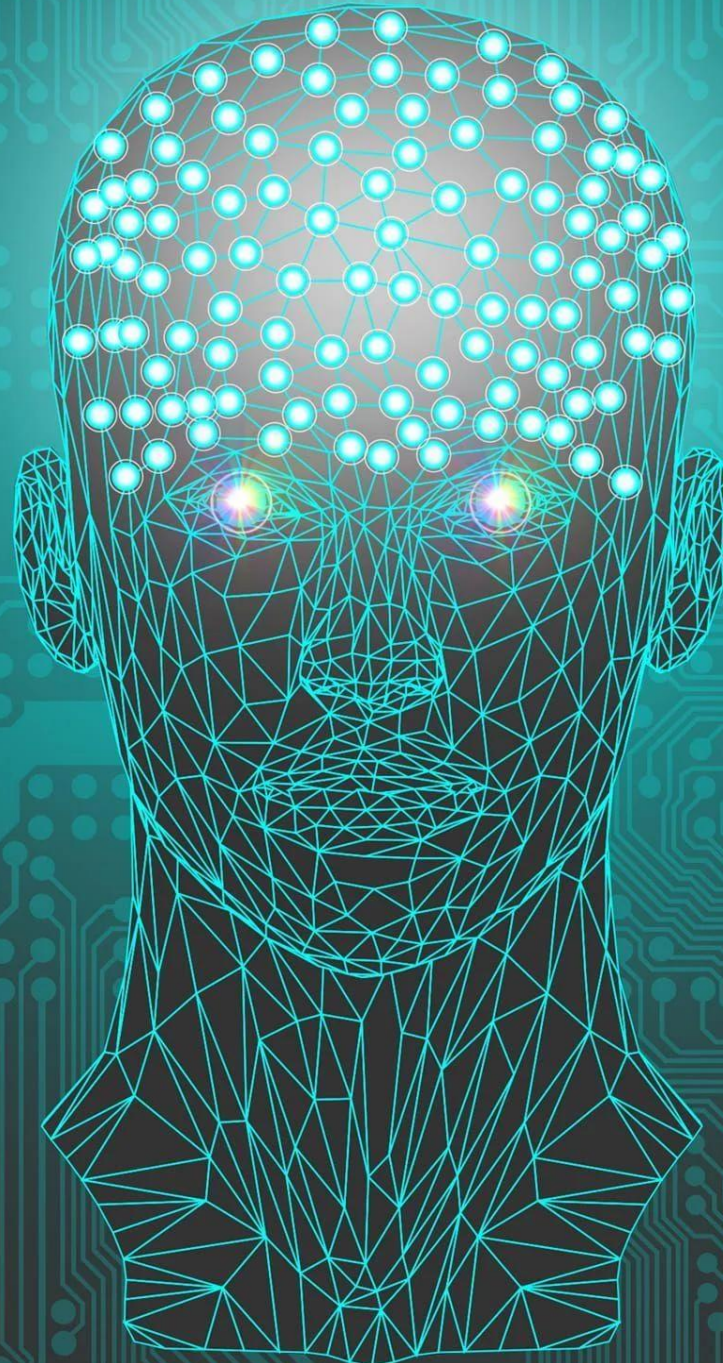
AI today is particularly adept at linguistic, semantic, generative and analytical activity especially across larger sets of “data”.

Digital Delegation

What activity could your organization “delegate” to AI to improve team performance? [We would call such delegated activities potential “*AI use cases*”]

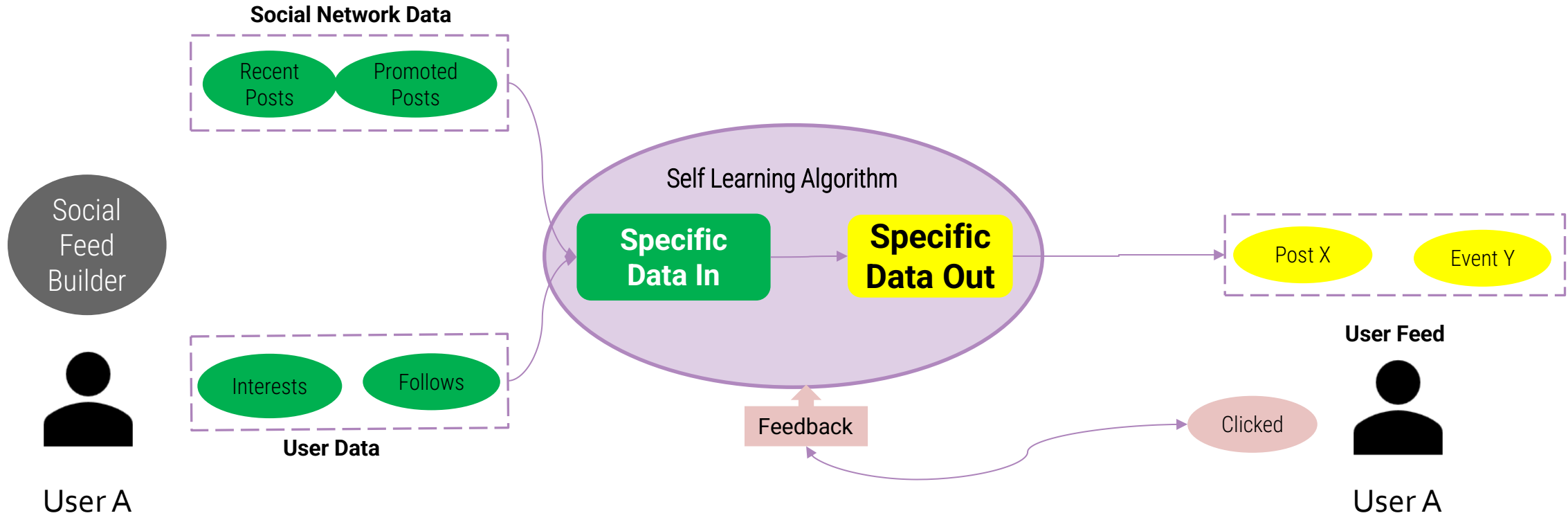


What can AI do?



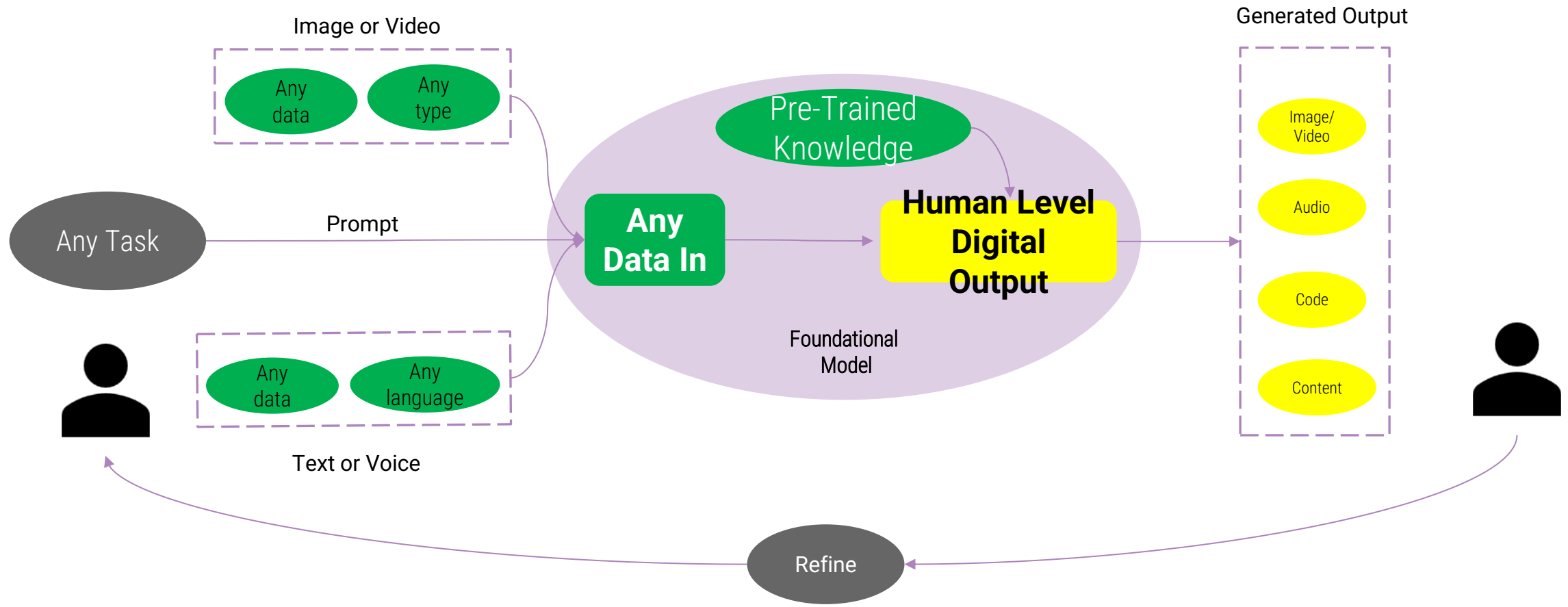
Understanding Machine Learning

Single Purpose or Narrow AI



Understanding Generative AI

Near Future Multi-Modal Models





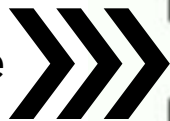
OpenAI Imagines Our AI Future

Stages of Artificial Intelligence

Level 1	Chatbots, AI with conversational language
Level 2	Reasoners, human-level problem solving
Level 3	Agents, systems that can take actions
Level 4	Innovators, AI that can aid in invention
Level 5	Organizations, AI that can do the work of an organization

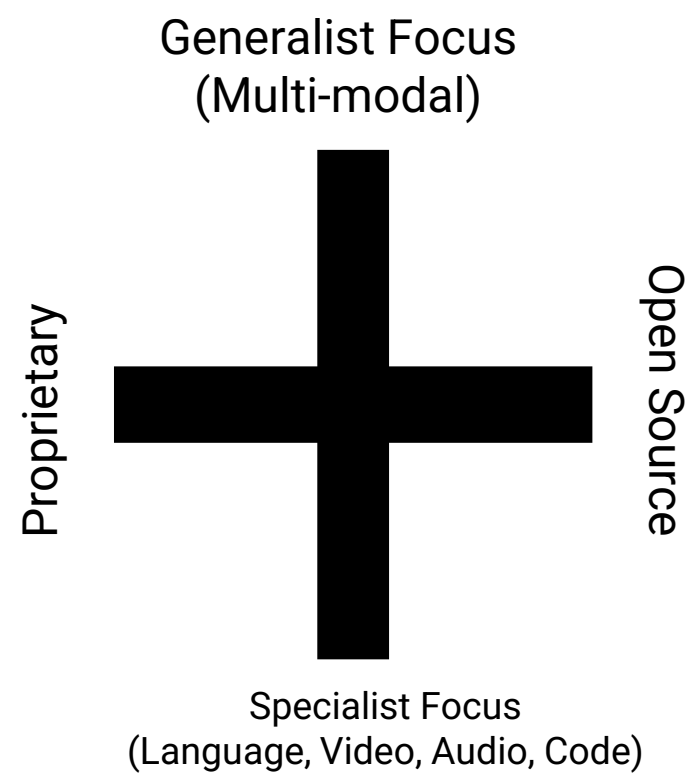
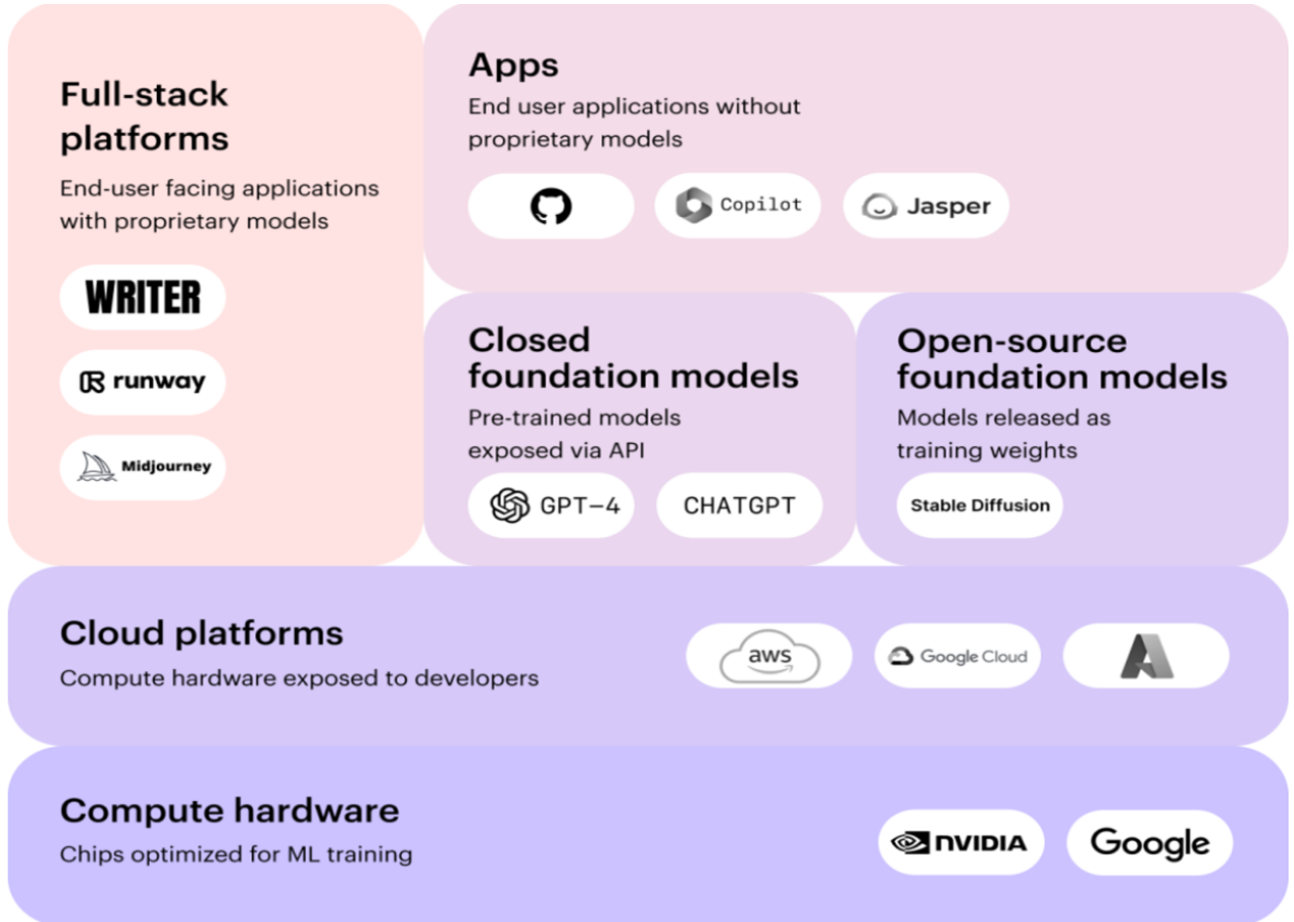
Source: Bloomberg reporting










We are here



Foundational Models

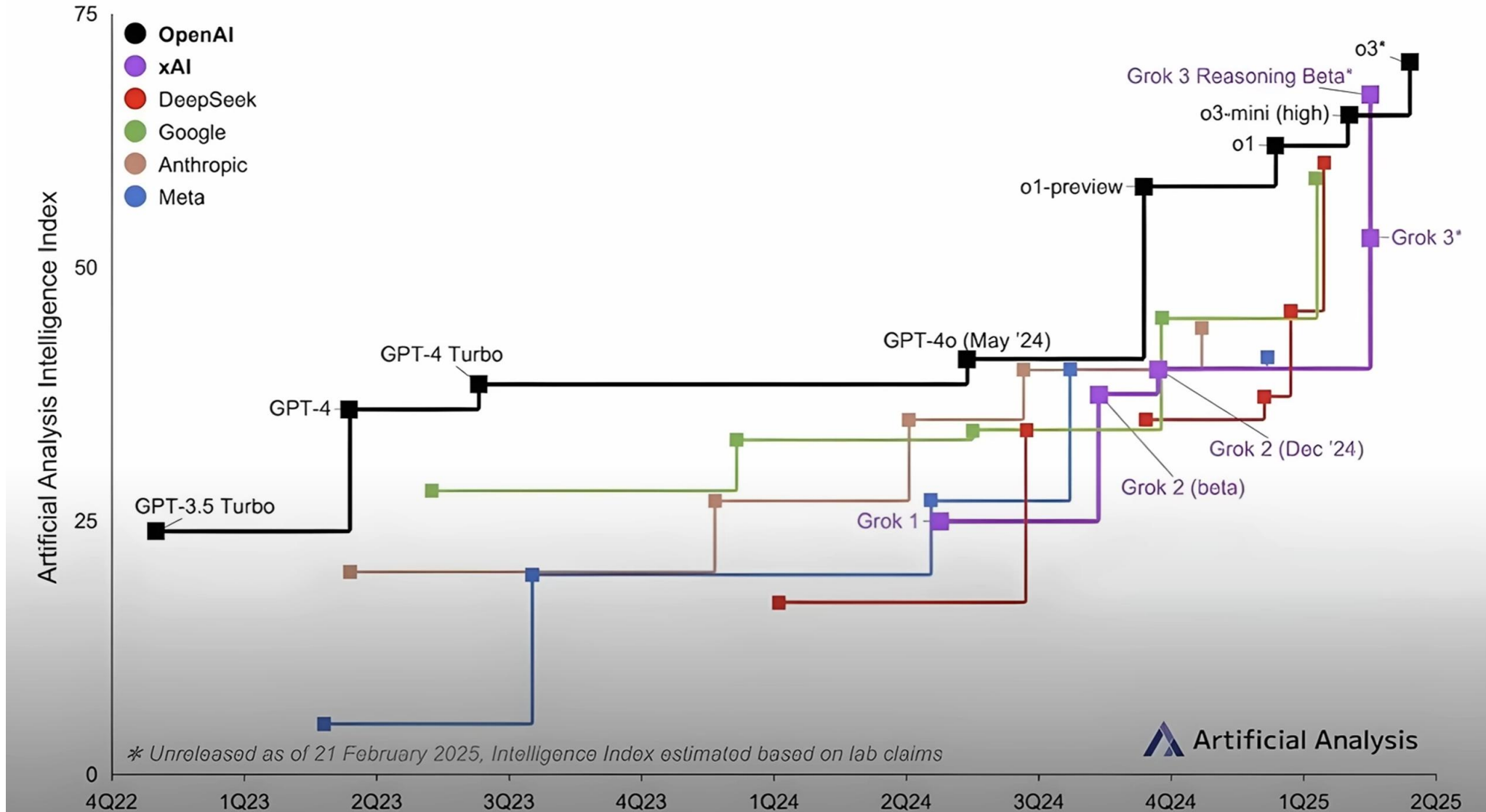
The "Engines" of Generative AI



- | | |
|--|---|
|  Gemini Family |  Meta Family |
|  OpenAI Family |  Hugging Face Family |
|  Snowflake Family |  Mistral Family |
|  Grok Family |  Deepseek Family |
|  Cohere Family | |

Which model performs the best?

Internet Sources Scraped February 2025



Comparing AI to Human Performance

Internet Sources Scraped December 2024

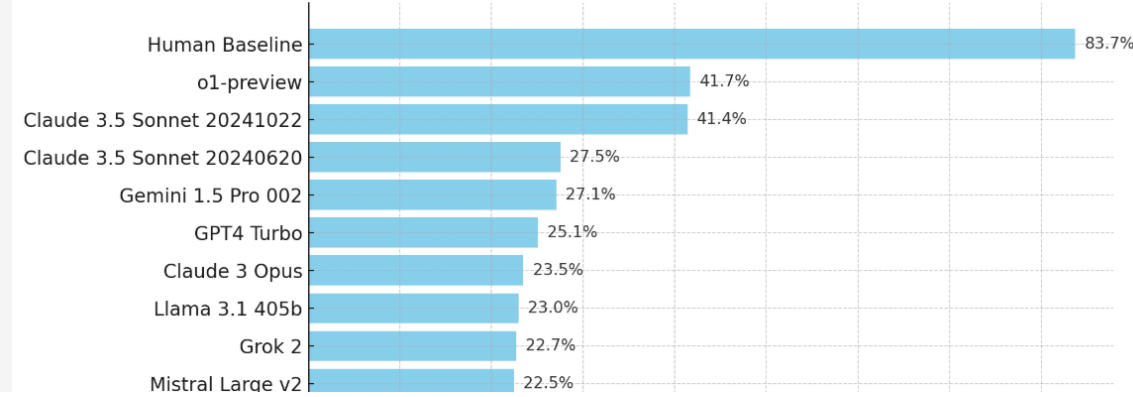
Key Score of 80% is generally equivalent to human level performance

	Claude 3.5 Sonnet (new)	Claude 3.5 Haiku	Claude 3.5 Sonnet	GPT-4o*	GPT-4o mini*	Gemini 1.5 Pro	Gemini 1.5 Flash
Graduate level reasoning <i>GPQA (Diamond)</i>	65.0% 0-shot CoT	41.6% 0-shot CoT	59.4% 0-shot CoT	53.6% 0-shot CoT	40.2% 0-shot CoT	59.1% 0-shot CoT	51.0% 0-shot CoT
Undergraduate level knowledge <i>MMLU Pro</i>	78.0% 0-shot CoT	65.0% 0-shot CoT	75.1% 0-shot CoT	—	—	75.8% 0-shot CoT	67.3% 0-shot CoT
Code <i>HumanEval</i>	93.7% 0-shot	88.1% 0-shot	92.0% 0-shot	90.2% 0-shot	87.2% 0-shot	—	—
Math problem-solving <i>MATH</i>	78.3% 0-shot CoT	69.2% 0-shot CoT	71.1% 0-shot CoT	76.6% 0-shot CoT	70.2% 0-shot CoT	86.5% 4-shot CoT	77.9% 4-shot CoT
High school math competition <i>AIME 2024</i>	16.0% 0-shot CoT	5.3% 0-shot CoT	9.6% 0-shot CoT	9.3% 0-shot CoT	—	—	—
Trustbit	70.4% 0-shot CoT	—	68.3% 0-shot CoT	69.1% 0-shot CoT	59.4% 0-shot CoT	65.9% 0-shot CoT	62.3% 0-shot CoT

SimpleBench

Evaluating Reasoning and Prompting

Model Scores (AVG@5) Compared to Human Baseline



model	code	crm	docs	integrate	marketing	reason	final 🏆	Cost	Speed
GPT-4o 🌐	90	95	100	90	82	75	89	1.21 €	1.50 rps
GPT-4 Turbo v5/2024-04-09 🌐	86	99	98	93	88	45	85	2.45 €	0.84 rps
Google Gemini Pro 1.5 0801 🌐	84	92	90	100	70	72	85	1.48 €	0.83 rps
GPT-4 v1/0314 🌐	90	88	98	52	88	50	78	7.04 €	1.31 rps
Claude 3.5 Sonnet 🌐	72	83	89	78	80	59	77	0.94 €	0.09 rps
GPT-4 v2/0613 🌐	90	83	95	52	88	50	76	7.04 €	2.16 rps
GPT-4 Turbo v4/0125-preview 🌐	66	97	100	71	75	45	76	2.45 €	0.84 rps
GPT-4o Mini 🌐	63	87	80	52	100	67	75	0.04 €	1.46 rps
Claude 3 Opus 🌐	69	88	100	53	76	59	74	4.69 €	0.41 rps
Meta Llama3.1 405B Instruct 🌐	81	93	92	55	75	46	74	2.39 €	1.16 rps
GPT-4 Turbo v3/1106-preview 🌐	66	75	98	52	88	62	73	2.46 €	0.68 rps
Mistral Large 123B v2/2407 🌐	68	79	68	75	75	71	73	0.86 €	1.02 rps

Real World Performance

Internet Sources Scraped December 2024

Use case : The model is asked to correct contract text; a fix that a lawyer might write and send to the opposing party for review. Contract types were Non-Disclosure Agreements (NDA), Data Processing Agreements (DPA), Master Service Agreements (MSA), Sales Agreements, and Employment Agreements.

Uses cases include : Taxable income calculation: Understanding the differences between accounting income and taxable income, including permanent and temporary differences; Tax rates: Applying the appropriate tax rates to calculate income tax expense; Deferred tax assets and liabilities: Recognizing and measuring deferred tax assets and liabilities arising from temporary differences.

ContractLaw Benchmark

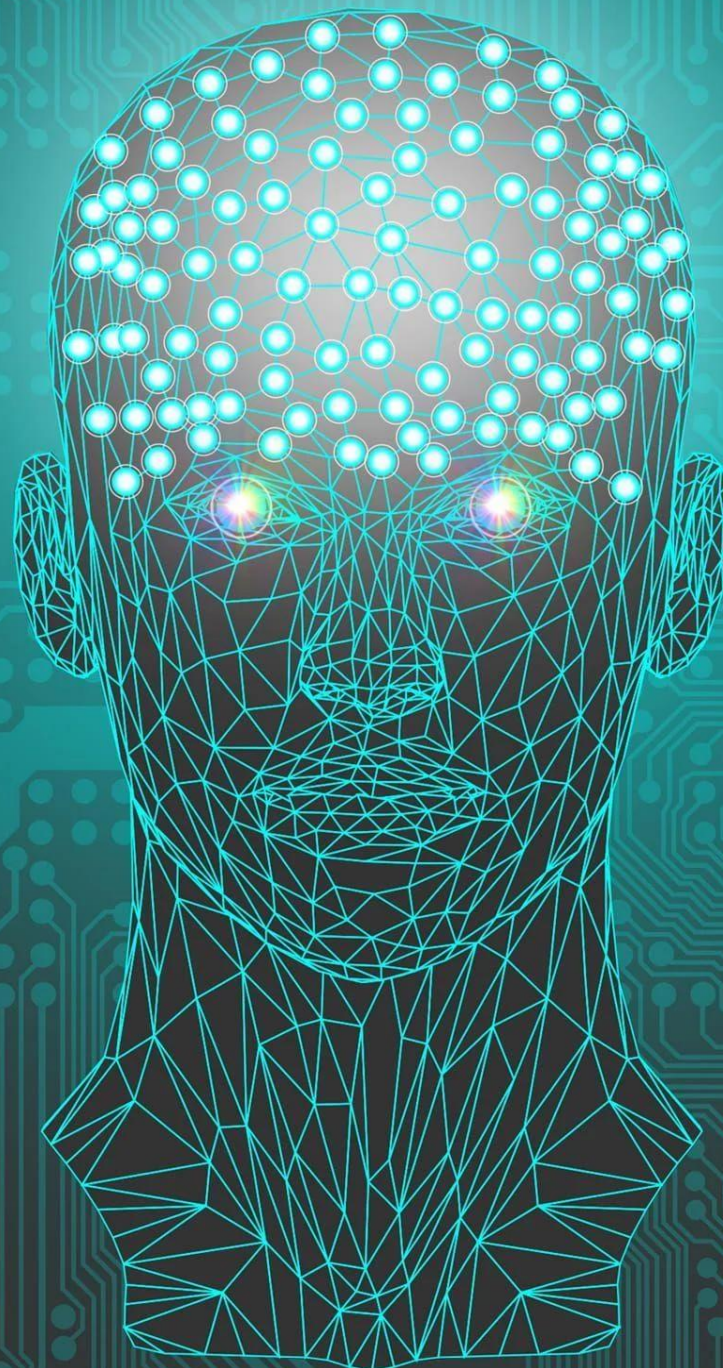
Model	Cost In / Out	Accuracy	Latency (s)
1 Llama 3.1 Instruct Turbo (405B)	\$3.50 / \$3.50	75.2 %	2.19 s
2 Claude 3 Opus	\$15.00 / \$75.00	74.0 %	5.97 s
3 o1 Mini	\$3.00 / \$12.00	72.8 %	4.01 s
4 GPT 4o Mini	\$0.15 / \$0.60	72.4 %	1.92 s
5 GPT 4	\$10.00 / \$30.00	71.8 %	3.26 s
6 o1 Preview	\$15.00 / \$60.00	69.0 %	12.83 s
7 Claude 3.5 Sonnet Latest	\$3.00 / \$15.00	68.7 %	2.28 s
8 Llama 3.1 Instruct Turbo (70B)	\$0.70 / \$0.70	68.6 %	4.74 s
9 Command R+	\$3.00 / \$15.00	68.2 %	1.17 s
10 Claude 3.5 Sonnet	\$3.00 / \$15.00	68.2 %	1.61 s
11 Gemini 1.5 Pro 001	\$1.25 / \$5.00	68.0 %	4.11 s
12 Claude 3 Sonnet	\$3.00 / \$15.00	67.6 %	3.03 s
13 Llama 3 (70B)	\$0.90 / \$0.90	66.8 %	2.92 s

TaxEval Benchmark

Model	Cost In / Out	Accuracy	Latency (s)
1 o1 Preview	\$15.00 / \$60.00	83.5 %	7.38 s
2 o1 Mini	\$3.00 / \$12.00	72.4 %	3.41 s
3 Claude 3.5 Sonnet Latest	\$3.00 / \$15.00	67.1 %	0.43 s
4 GPT 4o	\$2.50 / \$10.00	66.4 %	0.71 s
5 Claude 3.5 Sonnet	\$3.00 / \$15.00	65.7 %	0.65 s
6 GPT 4	\$10.00 / \$30.00	60.5 %	0.64 s
7 Claude 3 Opus	\$15.00 / \$75.00	58.4 %	2.06 s
8 Llama 3.1 Instruct Turbo (405B)	\$3.50 / \$3.50	57.8 %	0.81 s
9 GPT 4o Mini	\$0.15 / \$0.60	54.2 %	0.46 s
10 Llama 3.1 Instruct Turbo (70B)	\$0.70 / \$0.70	49.4 %	2.89 s
11 Claude 3 Sonnet	\$3.00 / \$15.00	48.0 %	0.81 s



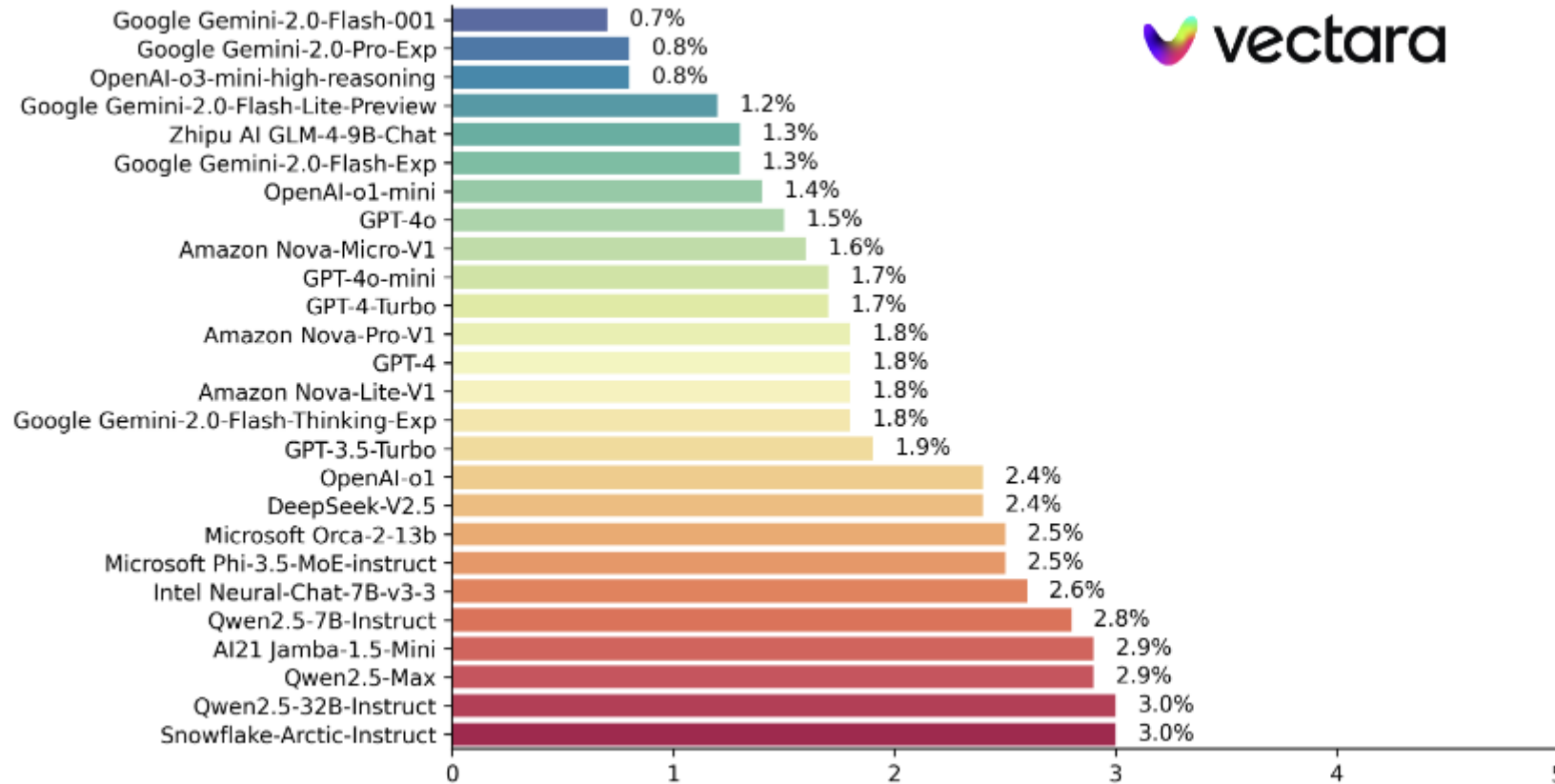
AI Safety, Risks and Limitations



AI Hallucinations

Can be reduced...but never eliminated....

Hallucination Rates for Top 25 LLMs



Last updated on February 5th, 2025

AI Hallucinations

Possible reasons and approaches to mitigation

Form of "Hallucination"	Probable Causes	Practical Mitigations
Factual	Training Data Used, Cut Off Date, Models not physically aware, Density	Use latest models, ensure AI applications include a "reflector agent".
Contextual or Logical	Chunking; Prompt Structure/Shots, Probabilistic	Leverage Large Context Windows, quality engineered prompting
Semantic	RAG "Lost in the Middle", Temperature	Curate knowledge base for RAG, lower temperature.

AI Hallucinations

A new form of software bug?

"Hallucination"	Software Bugs
Probabilistic in nature – cannot always be replicated or remediated.	Deterministic in nature – can be replicated and patched.
Not always self-evident; output may be plausible and more likely to lead to exposure to risk for longer period of time. Humans tend to expect 100% accuracy from technology.	More likely self-evident and easily identified. Negative business impacts does occur but more likely worked around or addressed after first occurrence.
(Much) higher levels of risk can be imagined when A is associated with human level agency.	Humans usually exist to address problems with automation ("Computer says No") risks.

AI Hallucinations

Does this mean AI should be “off the table”?

“The general reaction to language models among knowledge workers is one of denial. They grasp at the ever diminishing number of places where such models still struggle, rather than noticing the ever-growing range of tasks where they have reached or passed human level.”

“The economically and politically relevant comparison on most tasks is not whether the language model is better than the best human, it is whether they are better than the human who would otherwise do that task.”

— Avital Balwit, Rhodes Scholar and Chief of Staff to Anthropic CEO Dario Amodei



AI Concerns proliferate

AI dis-intermediation of human relationships



How will we be able to tell apart human and AI content or actions?



(The World ID AI Powered Identity Verification Orb)

Who is liable when AI causes harm?



What should we use (and not use) AI for?

What remains uniquely human?

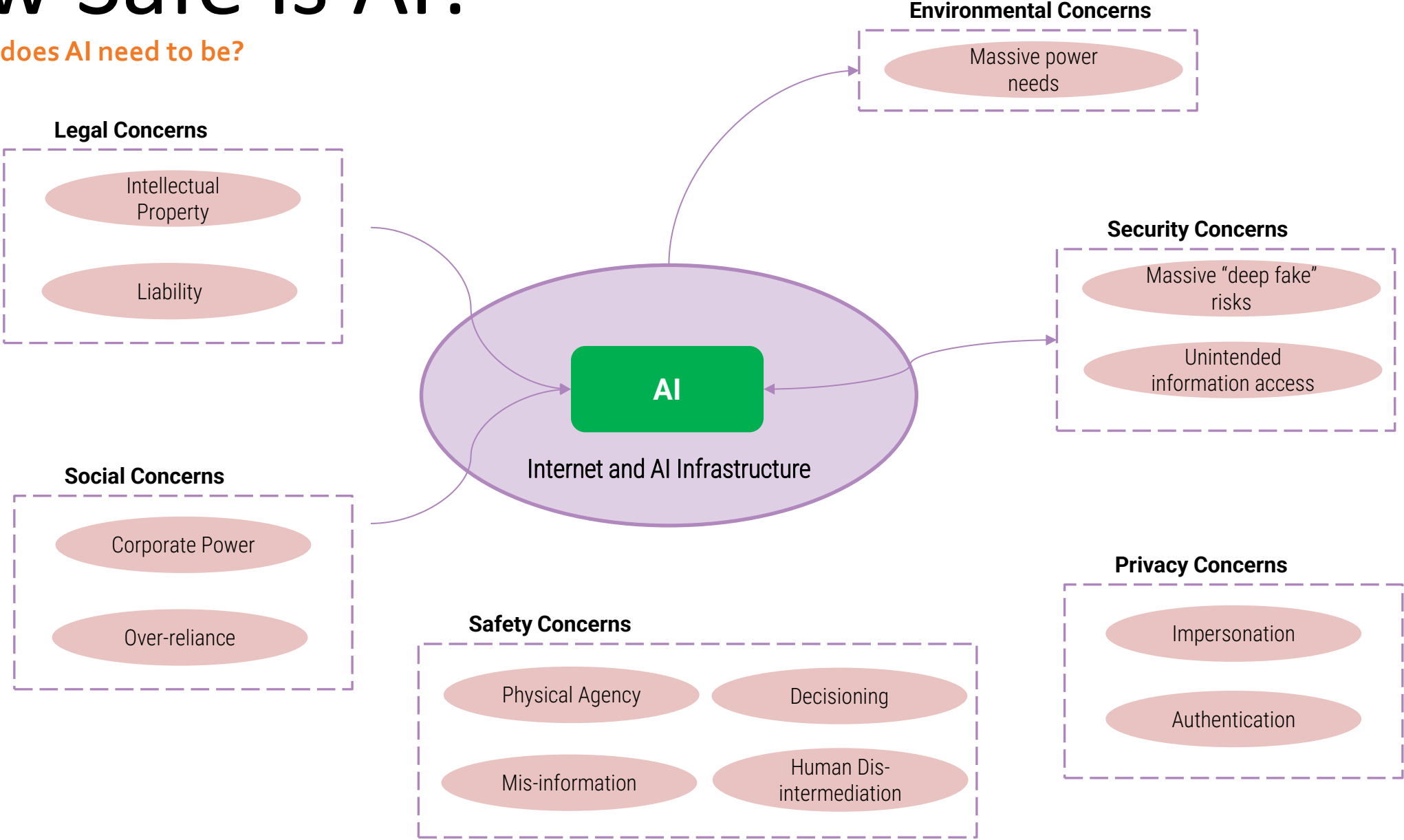
The question becomes what won't AI be able to achieve, or simulate, at or above human levels?

- Adaptability
- Common sense
- Compute efficiency
- Consciousness
- Creativity
- Curiosity
- Emotions
- Empathy
- Experience
- Ethics
- Imagination
- Instinct
- Intuition
- Love
- Morals
- Relationships
- Self-awareness
- Senses

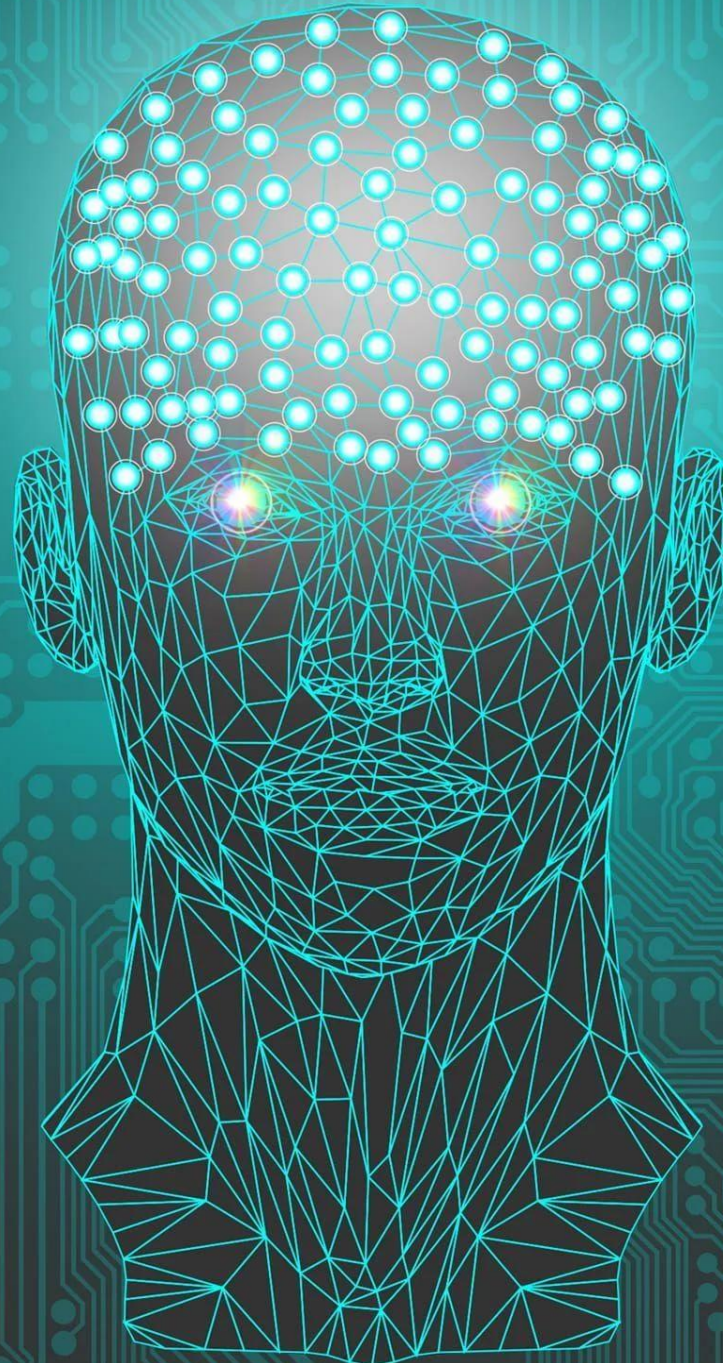


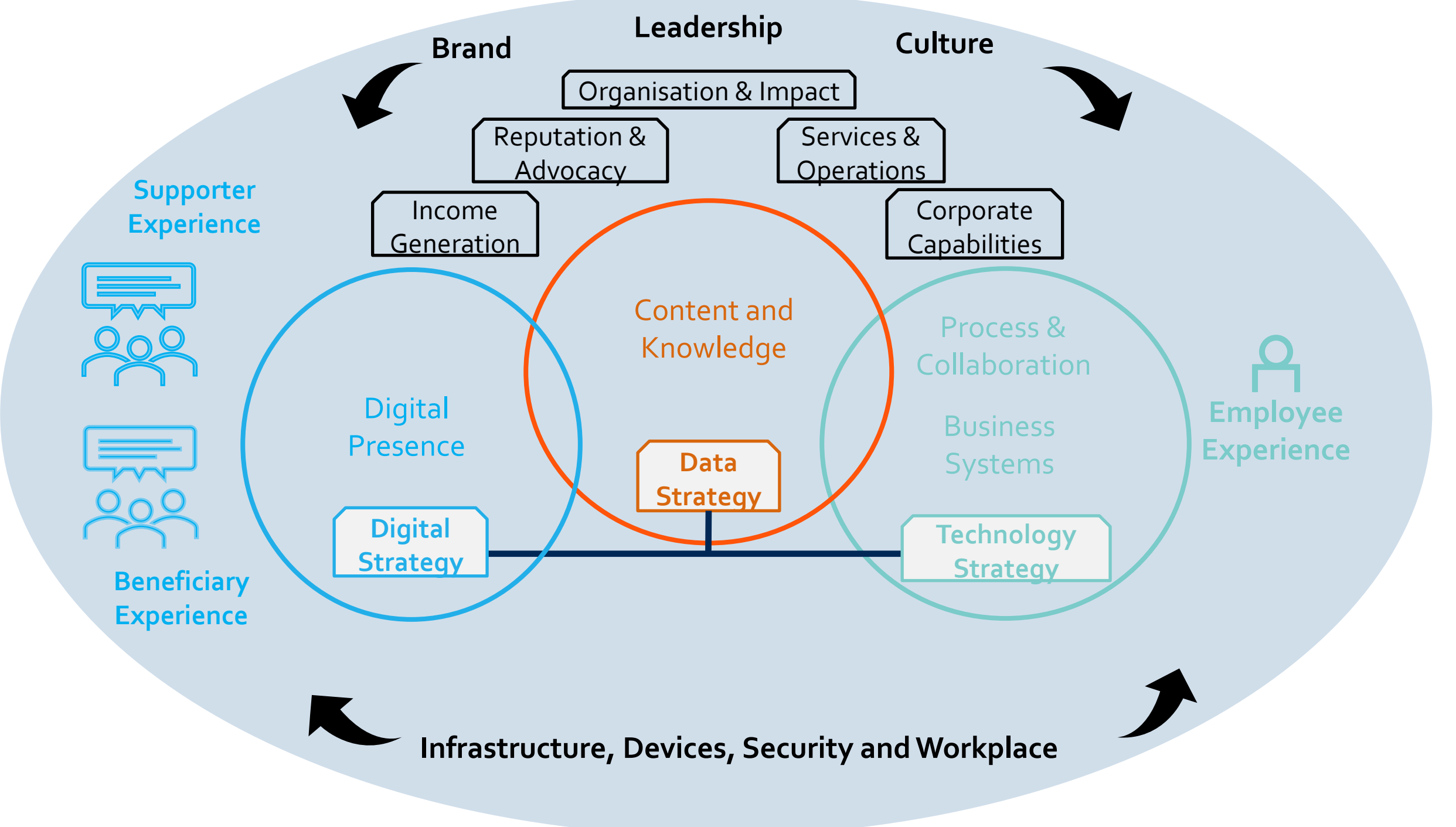
How Safe is AI?

How safe does AI need to be?



**Strategic use cases
for AI**





Claims made for the potential of AI

And how it would impact Marketing by the end of this decade

“It will mean that 95% of what marketers use agencies, strategists, and creative professionals for today will easily, nearly instantly and at almost no cost be handled by the AI — and the AI will likely be able to test the creative against real or synthetic customer focus groups for predicting results and optimizing. Again, all free, instant, and nearly perfect. Images, videos, campaign ideas? No problem.”

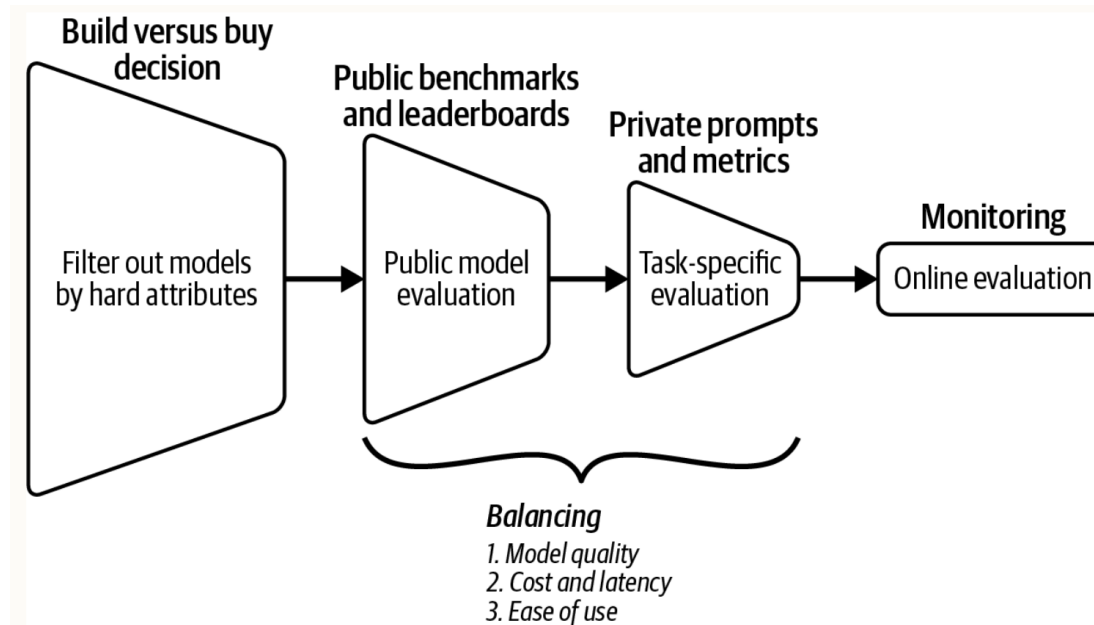
— Sam Altman, CEO, OpenAI



Evaluation Framework

For AI employed in your organisation

- **Factual Consistency** (in a given context) - (example Benchmark TruthfulQA)
- **Instruction Following** (system/user prompts) - (example IFEval/INFOBENCH)
- **Persona Adoption** (guardrails/memory) – (example RoleLLM/Character Eval)



- **Your Values** (in a given context) - (you need examples of good practice)
- **Your Outputs** (from real world cases) - (you need example work outputs)
- **Your Experience** (in a given function or service) – (you need user feedback)

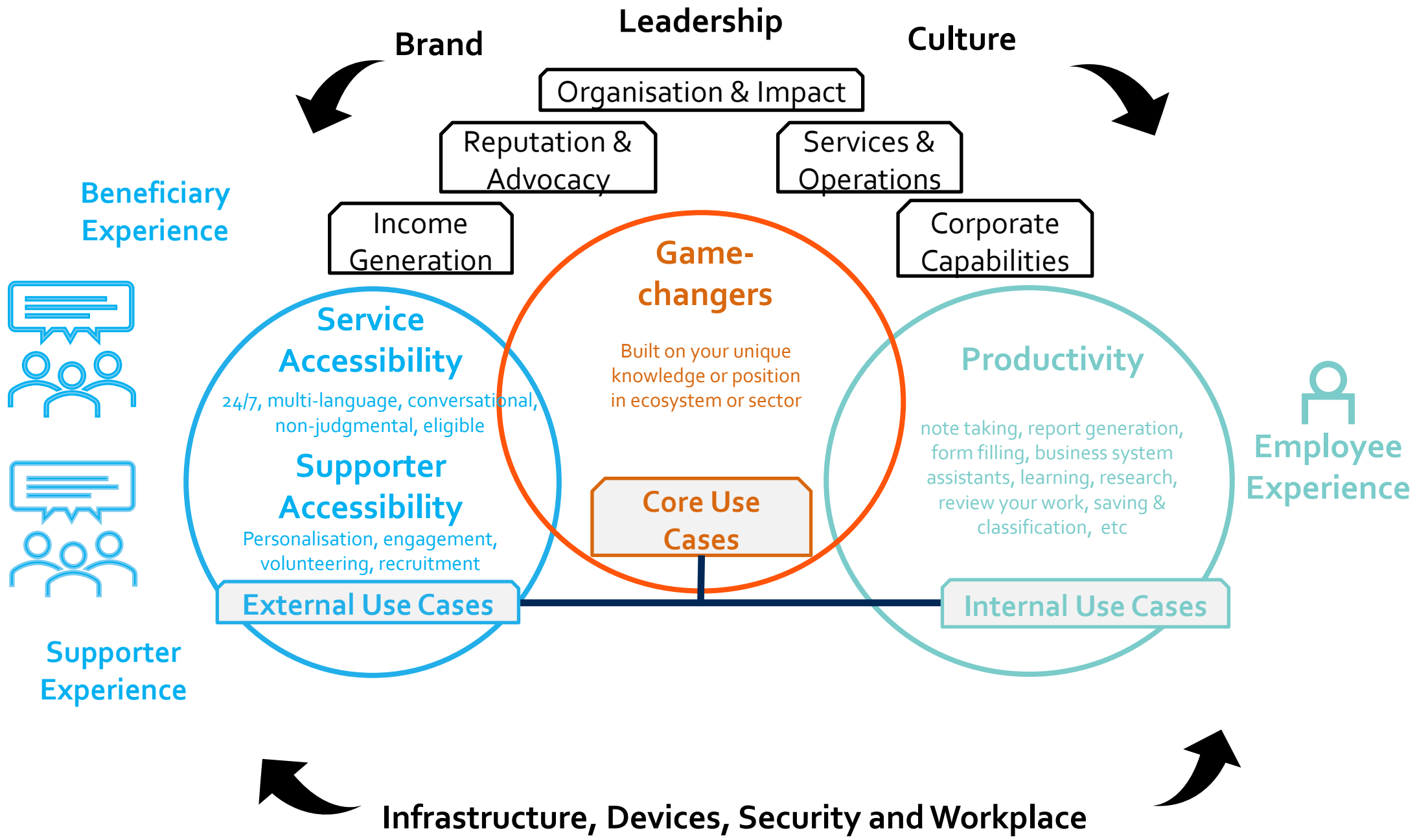


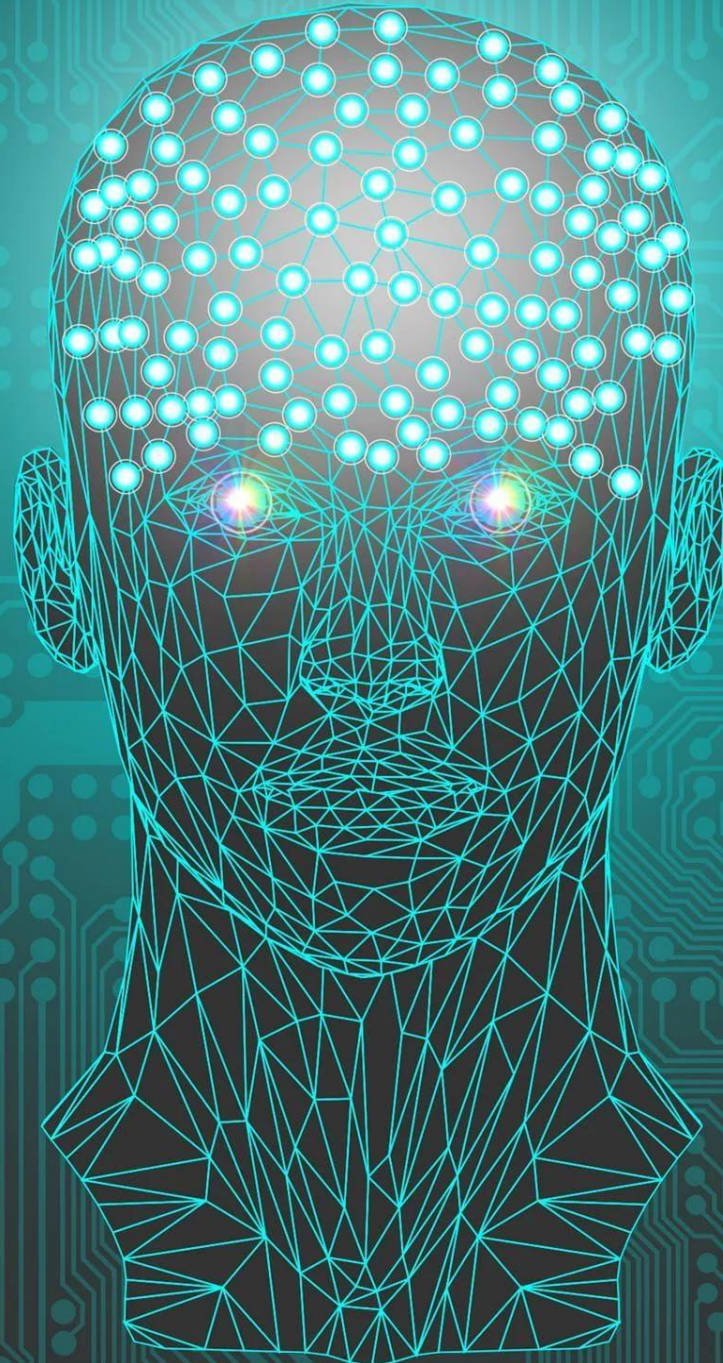
Table Discussions and



The most valuable ways AI could your organisation
in the next 3 years.

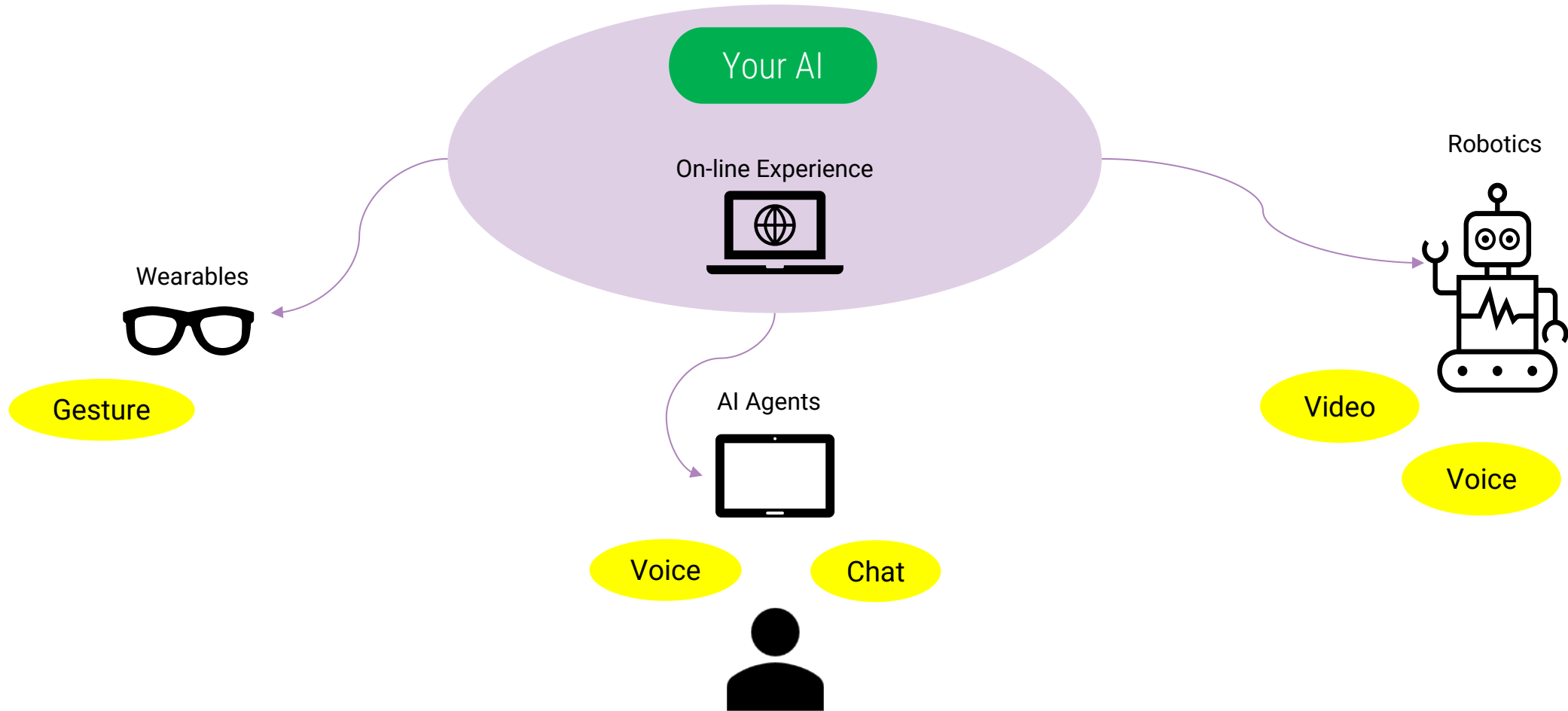
The AI marketplace for third sector organisations

The AI market



The AI Interface

On-line, off-line, devices and robotics



In 2030 will anyone be using a keyboard to enter data or code?
Will people browse websites or have to navigate business systems as we do today?

Consumer AI Market Today

A new market of emergent AI Tools



Writing & Editing



AI-powered writing tools generate content, edit grammar, and even rephrase sentences for better readability.

Design & Creative



AI tools that assist with graphic design, video editing, and content creation, enabling more efficient and innovative creative processes.

Customer Service & Support



AI-powered chatbots, automated responses, and sentiment analysis are enhancing customer interactions and streamlining support processes.

Growth & Marketing



AI that helps in optimizing marketing campaigns, analyzing customer data, and personalizing content to boost engagement and conversion rates.

Workflow Automation



AI that automates HR tasks, accounting processes, and document management, reducing manual effort and errors.

Technology & IT

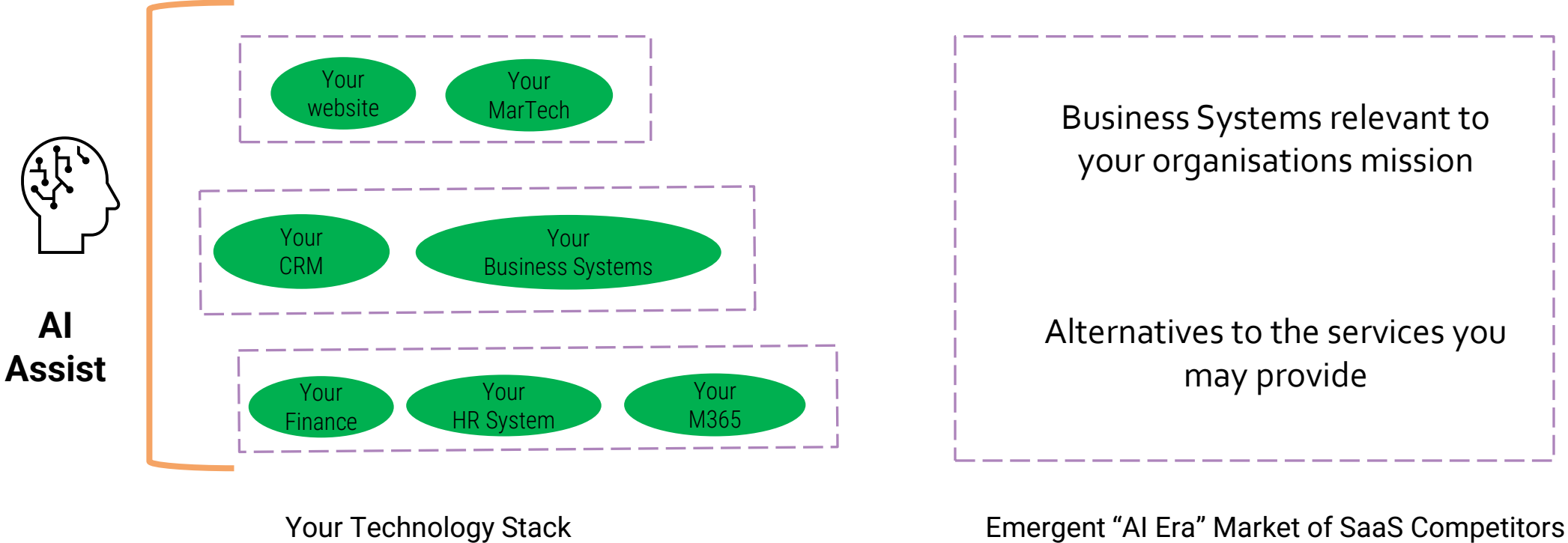


Intelligent AI website builders, e-commerce solutions, code assistants, and more to optimize your digital operations.

Enterprise Solutions Market

AI Capabilities for existing solutions

Emergent “AI first” competitors



Larger vendors in your technology stack will likely release AI capabilities to meet common customer use cases dependant on content or data in that particular solution.

Microsoft Copilots

M365 Copilot

Bing



Better Q&A and task completion

Edge



Better interaction with web content

Word



Better reading and writing assistance

Outlook



Better e-mail management

Excel



Better data analysis

PowerPoint



Better presentations

Teams



Better Meetings

Business Chat



Better knowledge management

Windows Copilot



Better interaction with OS, apps, and files

Copilots for Web

Copilots for Productivity

Copilot for Everyday

Dynamics Copilot



Better sales and customer support

Copilots for Business

Fabric Copilot



Better data analytics and business intelligence

Copilots for Analytics

Security Copilot



Better threat detection, identification, and mitigation

Copilot for Security

GitHub Copilot



Better code development

Copilot for Development

Power Platform Copilot

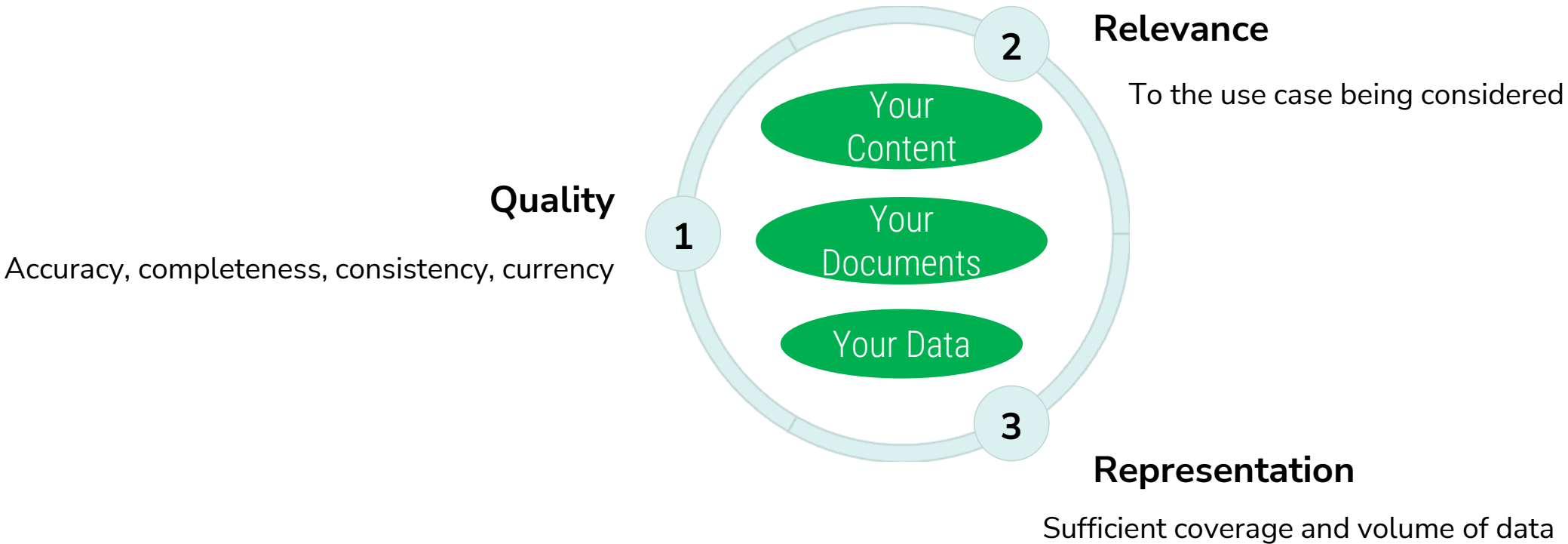


Better creation of apps, workflows, and agents

Copilot for Low/No Code Development

Your Organisational Knowledge

And the data you need to “fit” to each AI use case.



(Data fit is the alignment between the data you use and the AI use case you want

to solve)

LAYER 2 INTELLIGENCE LAYER

Frameworks

LangChain, LlamaIndex, Haystack, DSPy, PydanticAI, Llama, Langbase, TaskWeaver, compoo, AI Suite, Promptflow, LLMSack

Knowledge Engines

Pinecone, Chzoma, Epsilon, milvus, drant, MongoDB, supabase, contextual.ai

FalcorDB, Superduper

Specialized Coding Models

Claude, Open2.5, poolside

LAYER 1 INFRASTRUCTURE LAYER

AI Workspaces

Daytona, RUNLOOP, E2B

Model Access & Deployment

OpenAI, ANTHROPIC, GROQ, AIZI labs, cohere, Hugging Face, Cartesia, Fireworks AI, together.ai, Replicate, SymbioNEX, OpenRouter

Cloud Providers

AWS, Koyeb, Hyperbolic, CoreWeave, NEBIUS

LAYER 5 AGENT CONSUMER LAYER

Autonomous Agents

Cognition, Bolt.new, Coworked

Assistive Agents

GitHub Copilot, Continue, Sourcegraph, CURSOR, bobnine, supermaven, windsurf, HEX, bloop.

Specialized Agents

CodeRabbit, qodo, Ellipsis, codelflash, Superflex, /codemod., Codegen

LAYER 3 ENGINEERING LAYER

Training & Fine-Tuning

LAMINI, Prebase, Modal, Julius, Cohere, codalysphere, OpenAI

Tools

Relevance AI, Sourcegraph, PromptLayer, gretel, MOSTLY.AI, TOMIC, JigsawStack

Testing & Quality Assurance

Adaline, LangSmith, Langfuse, arize, Weights & Biases, AgentOps.ai, Confident AI, ContentDR, braintrust

LAYER 4 OBSERVABILITY AND GOVERNANCE LAYER

Development Pipeline

portkey, baseten, LangServe, STACK

Evaluation & Monitoring

Pydantic Logfire, Cleanlab, Patronus AI, LOC10, traceloop, WHYLABS, OpenLLMify, LangWatch

Risk & Compliance

allina, Guardrails AI, LAKERA, Socket

Security & Access Control

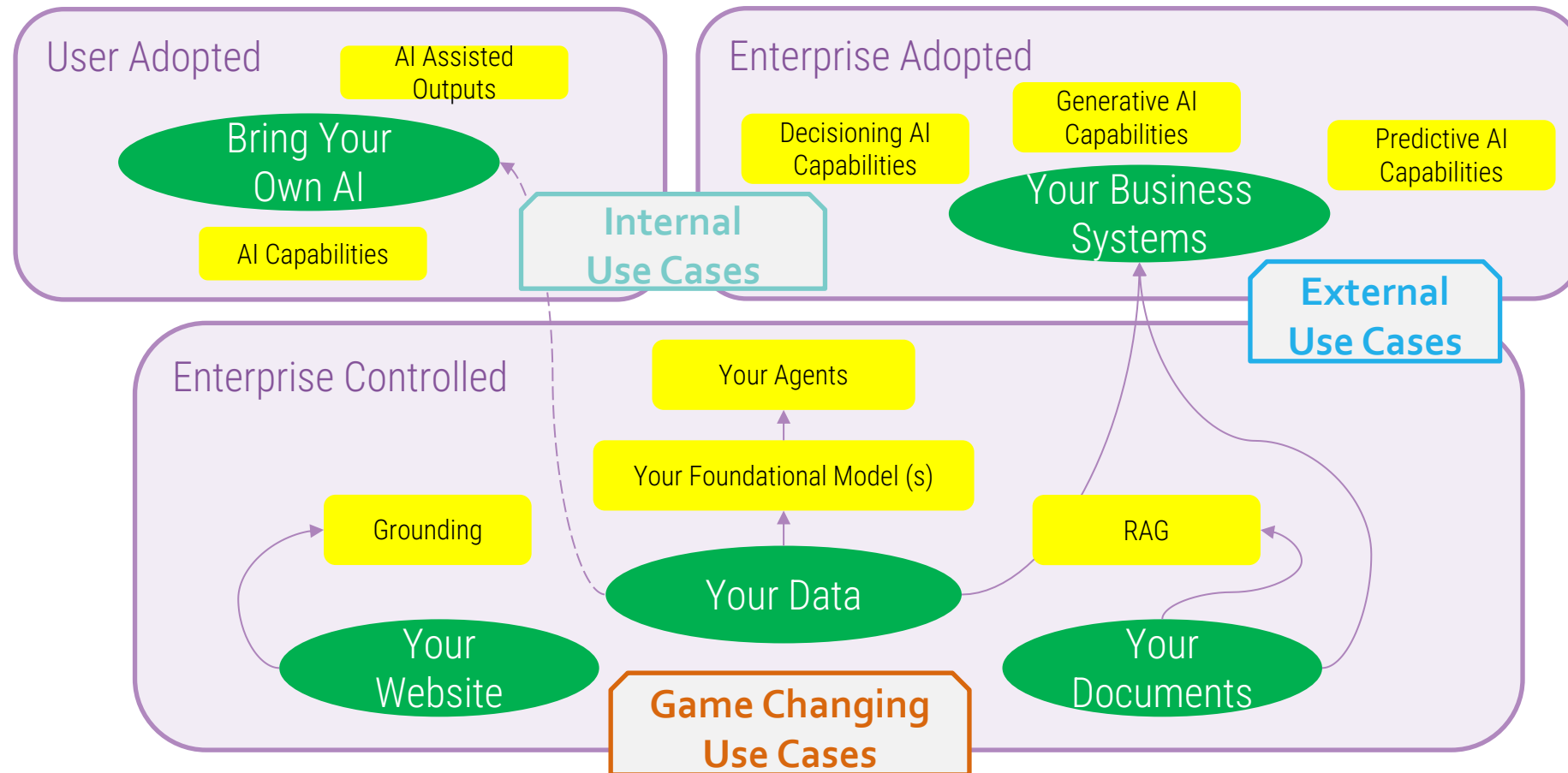
martian

Tech Stack for “Game Changing” AI

Highly Fluid, Rapidly Evolving Emergent Market

Your Emergent AI Tech Stack

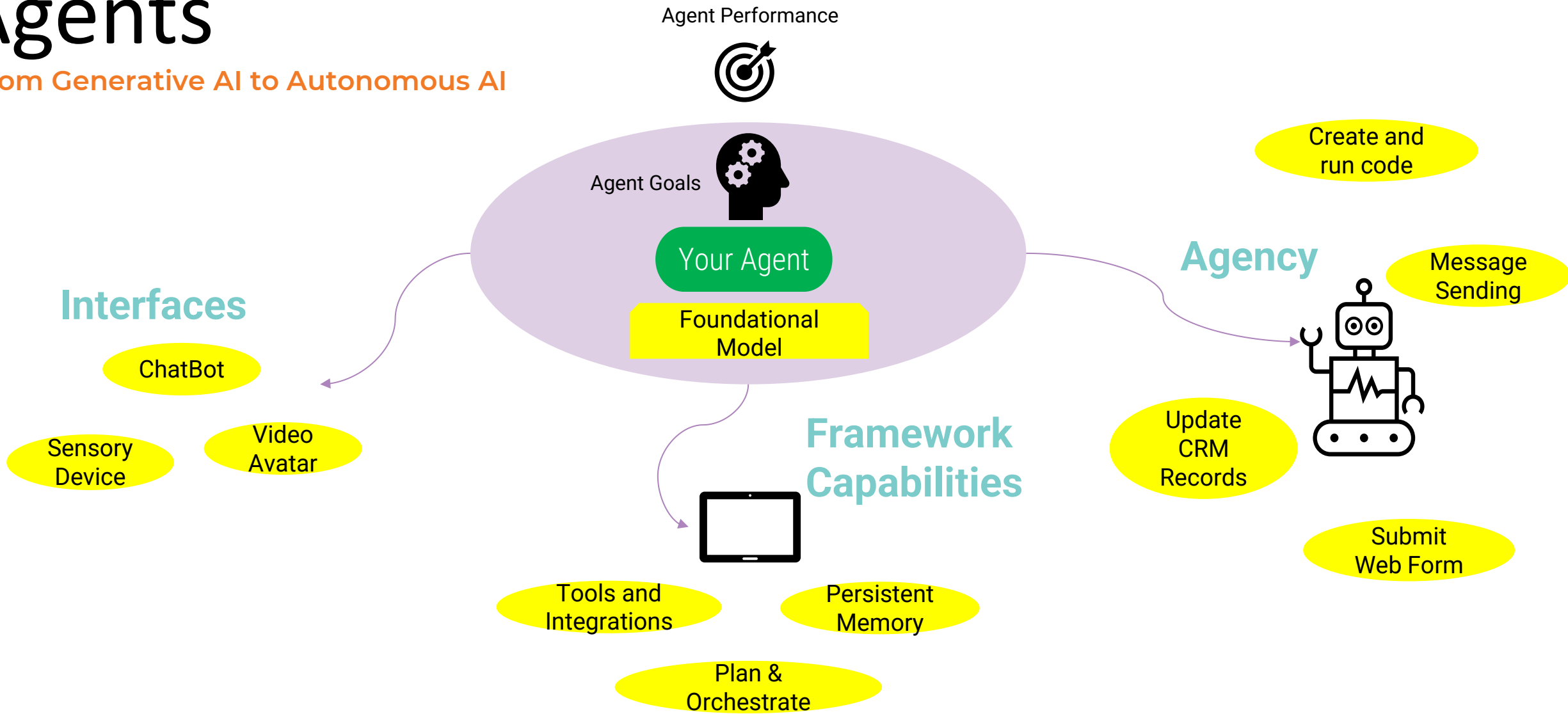
AI Everywhere – as a product, a product capability or a free internet tool.



Your quality of your knowledge (data and content) will be key to all AI use in your organisation.

Agents

From Generative AI to Autonomous AI



In the future what could you delegate to an AI Agent in your organisation?

The AI Agents Market

AI Low Code – For Your Digital, Data or Tech Teams



Your use case data sits in multiple clouds or does not sit in Microsoft or Salesforce

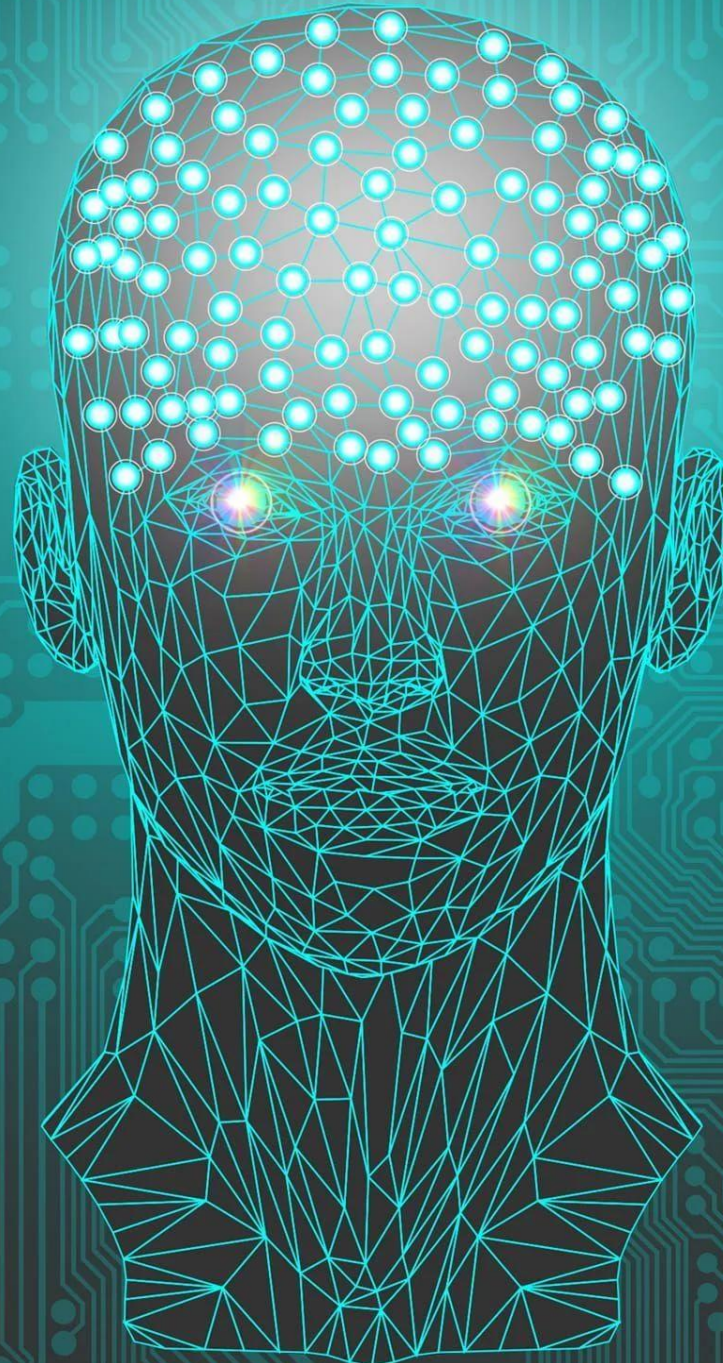
Your use case data is centred in the Microsoft cloud

Your use case data is largely in Salesforce

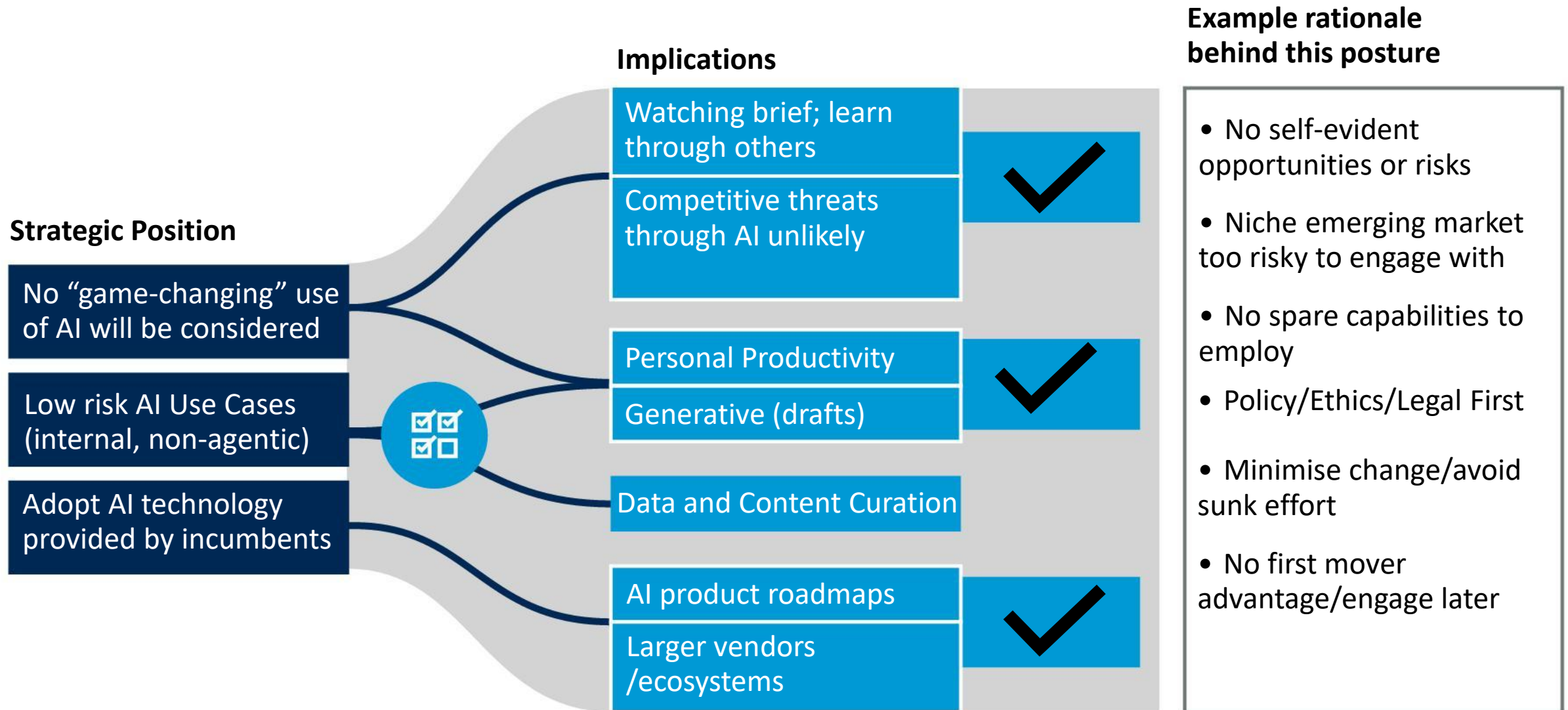
Development Frameworks (for implementor use)



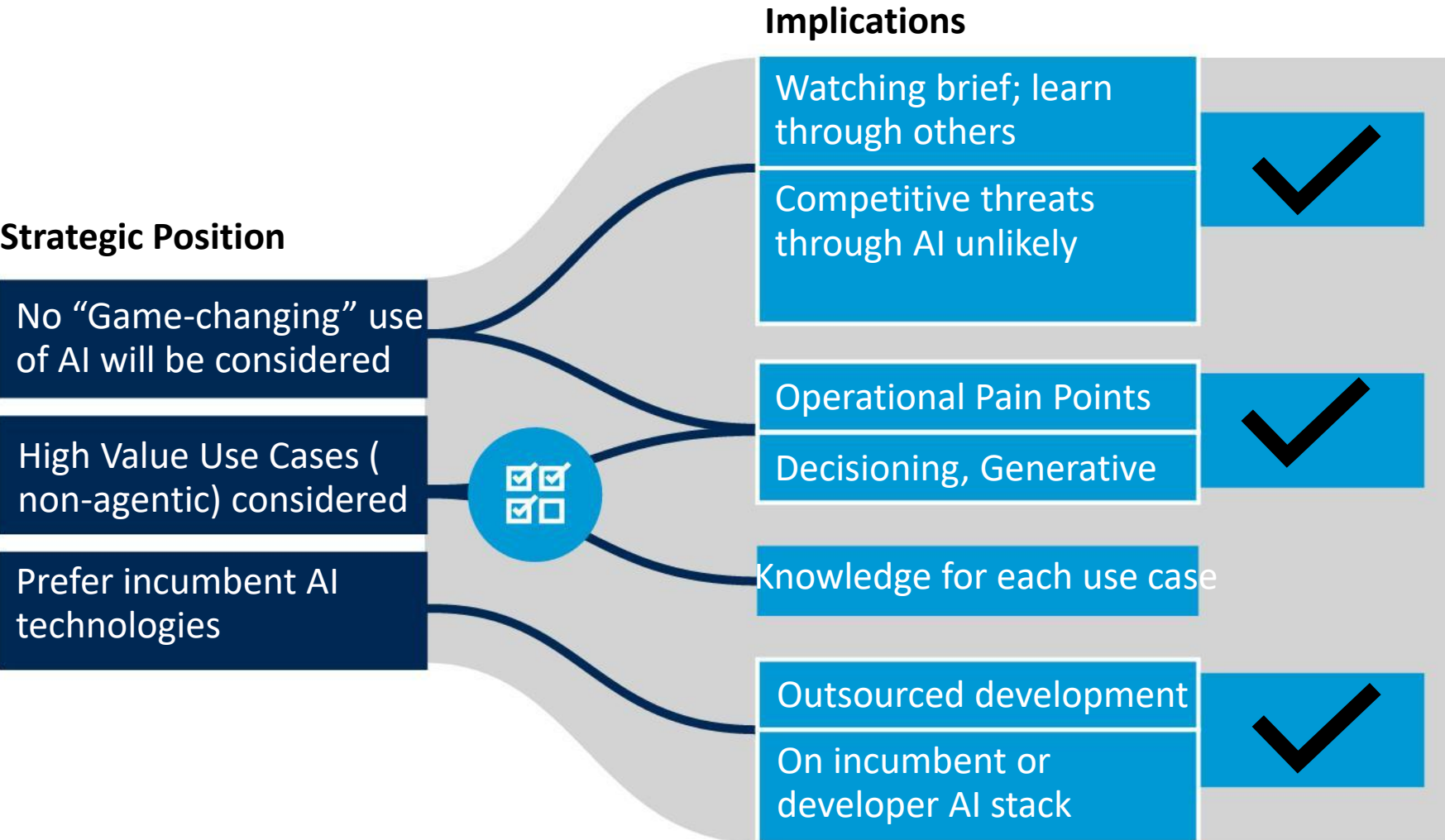
**How could
organisations
respond to the
opportunities and
risks associated
with AI?**



Slow Follower



Fast Follower



Example rationale behind this posture

- Self-evident opportunities, low competitive risk
- Willingness to "dip toe" into emergent AI market
- Spare capabilities to employ/talent to entertain
- Policy/Ethics/Legal in parallel
- Willing to invest to learn and gain some operational benefits.
- Recognition of potential of AI/need to start journey now

Third Sector Innovator

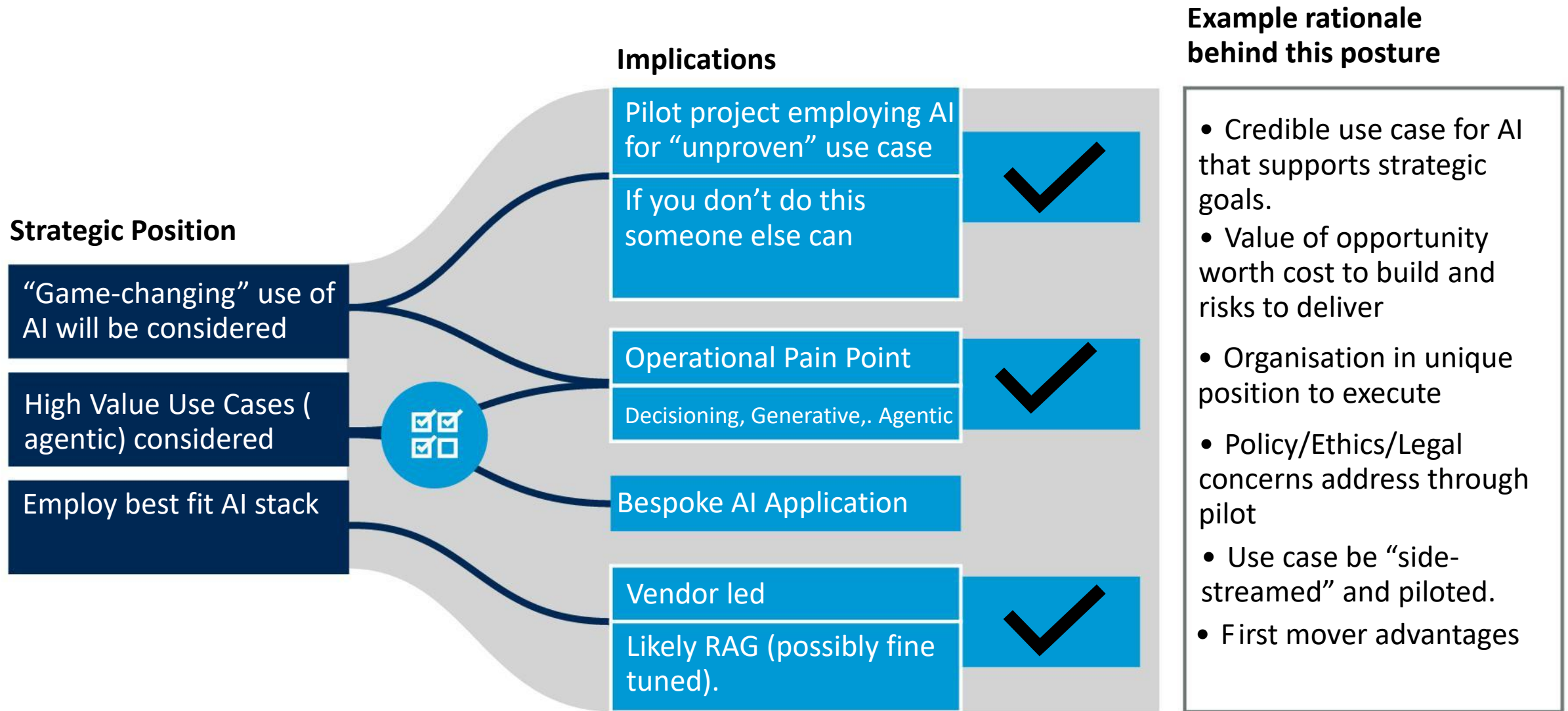


Table Discussions



What solutions have been employed to date, what has been learnt to date and how might you build AI capabilities going forward?



Discussion and feedback

Event feedback

Please use the QR code to view and complete the online feedback form.





Thanks & Goodbye!

Upcoming events...

Strategic AI opportunities for non-profits
26 March - Zoom event

**Information security, data protection and
cake**
11 June, 2pm – 5pm, RCN, London

www.adaptaconsulting.co.uk/upcoming-events



help@adaptaconsulting.co.uk



www.adaptaconsulting.co.uk



Adapta Consulting, 5 St John's Lane, London, EC1M 4BH



020 4558 8070

We hope you find this presentation enjoyable and thought-provoking. Please note that this recording is provided for general information purposes only and does not constitute professional advice. No user should act on the basis of any material contained in the recording or any of its supporting materials without obtaining proper professional advice specific to their situation.

Adapta has made reasonable efforts to ensure that the information provided is accurate and reliable, however no warranty is given regarding the accuracy or reliability of such information. All content is subject to change at any time and without notice.

The presentation may include references to specific products or services and/or links to other resources and websites. These references and links are provided for your convenience only and do not signify that Adapta endorses, approves or makes any representation or claim regarding the accuracy, copyright, compliance, legality, or any other aspects of the products, services resources or websites to which reference is made.

This recording may contain confidential and/or proprietary information, and must not be re-used or disclosed to third parties without the prior written approval of Adapta Consulting LLP. © Adapta Consulting LLP 2025.

If you would like further information or any advice regarding your own specific issues, then please do contact the Adapta team at hello@adaptaconsulting.co.uk