Ethical & Effective AI Adoption

Practical Considerations

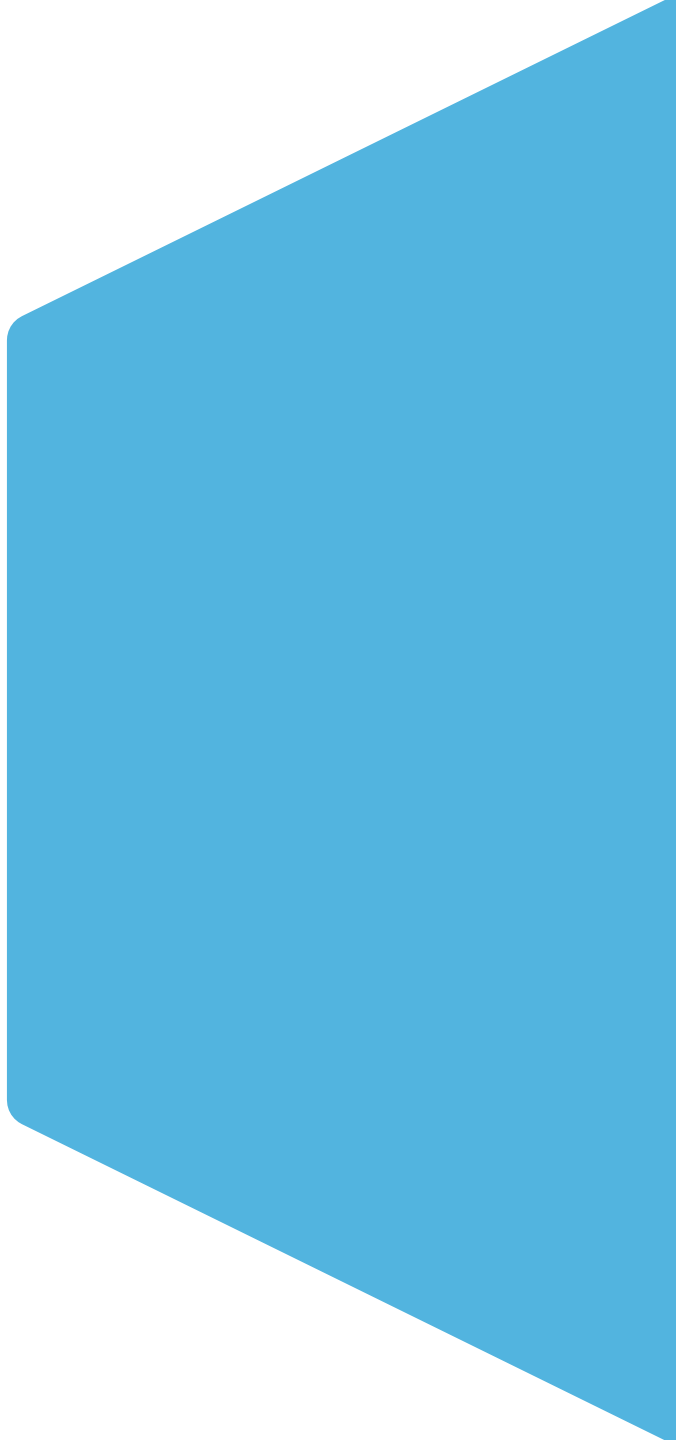
CASTLEBRIDGE

Fools rush in where angels fear to tread

Alexander Pope

CASTLEBRIDGE

What do we mean by AI?



AI has been used by vets since 1784

CASTLEBRIDGE

The Problem of Branding – Lots of things are called “AI”

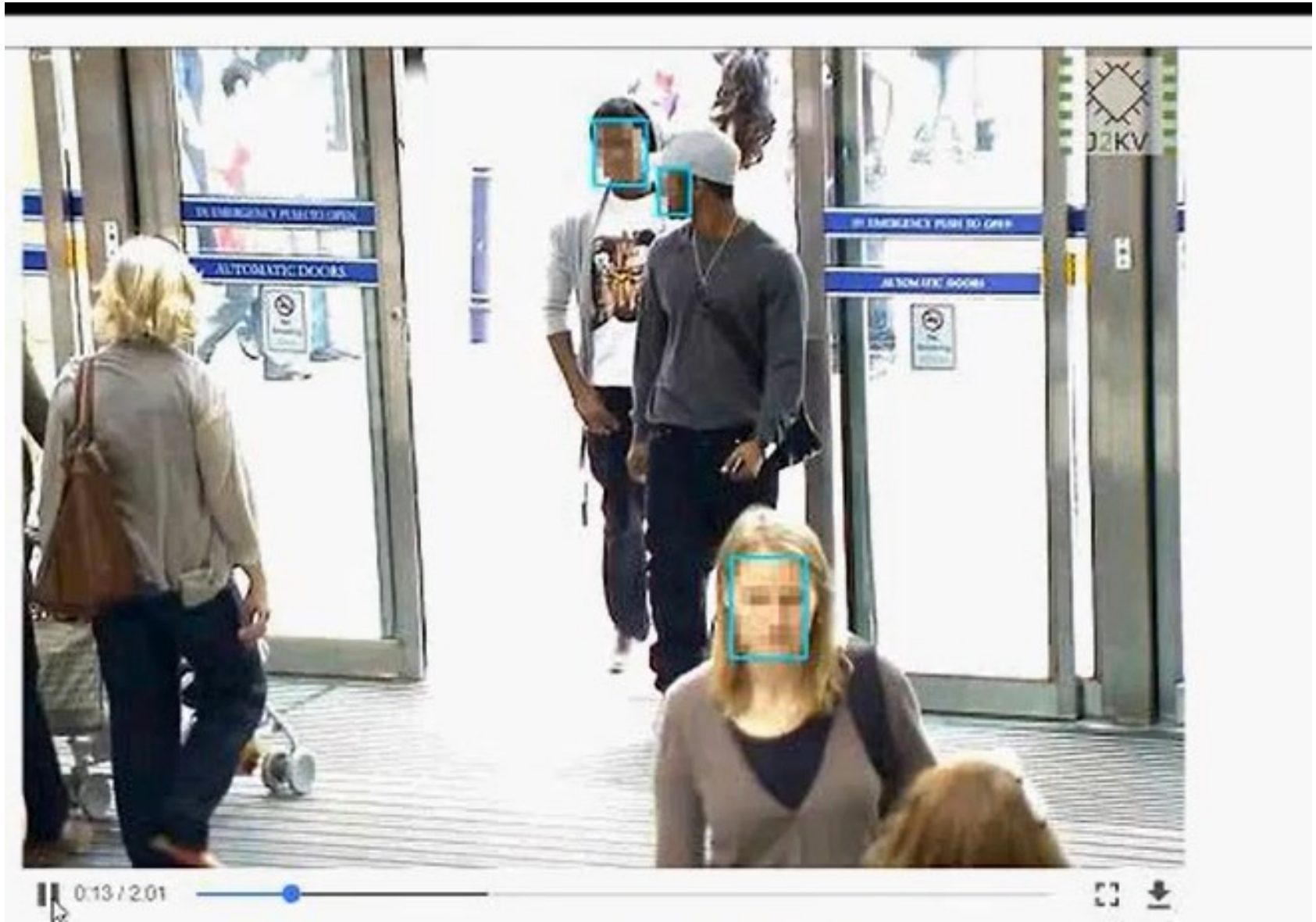
- **Robotic Process Automation**
  - Automation of repetitive tasks
  - Executed using scripted processes
- **Machine Learning**
  - Development of automated processing through either
    - Supervised Learning (humans tag information)
    - Unsupervised Learning (computer processes infer relationships)
- **NLP (Natural Language Processing)**
  - Processing of text to recognise words and patterns.
  - Useful for email filters, voice assistants, Autocorrect, Document summaries, information extraction
- **Artificial Intelligence**
  - “All of the above”

CASTLEBRIDGE

Example: Optical Character Recognition

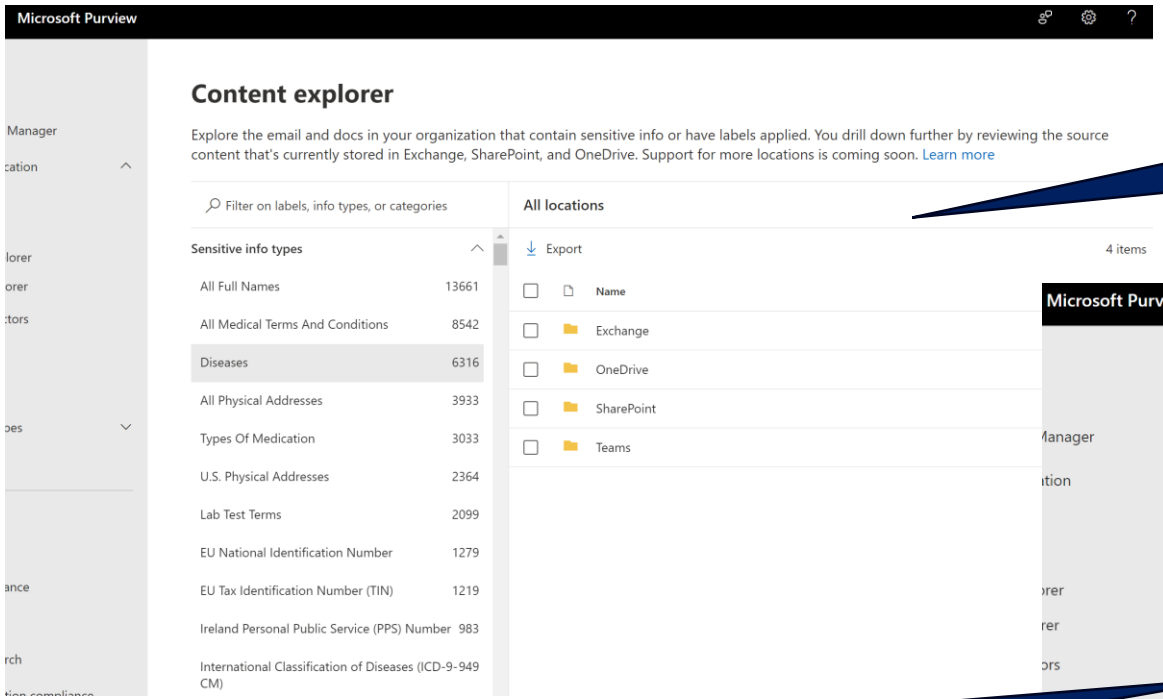
- Application of machine learning
- Uses statistical model of previously scanned documents to identify characters in images and non-text based
- Uses RPA, Machine Learning, and NLP
- Can make images accessible to other ML processes like automated data classification / categorisation processes

Example: Facial Detection (not the same as Recognition)



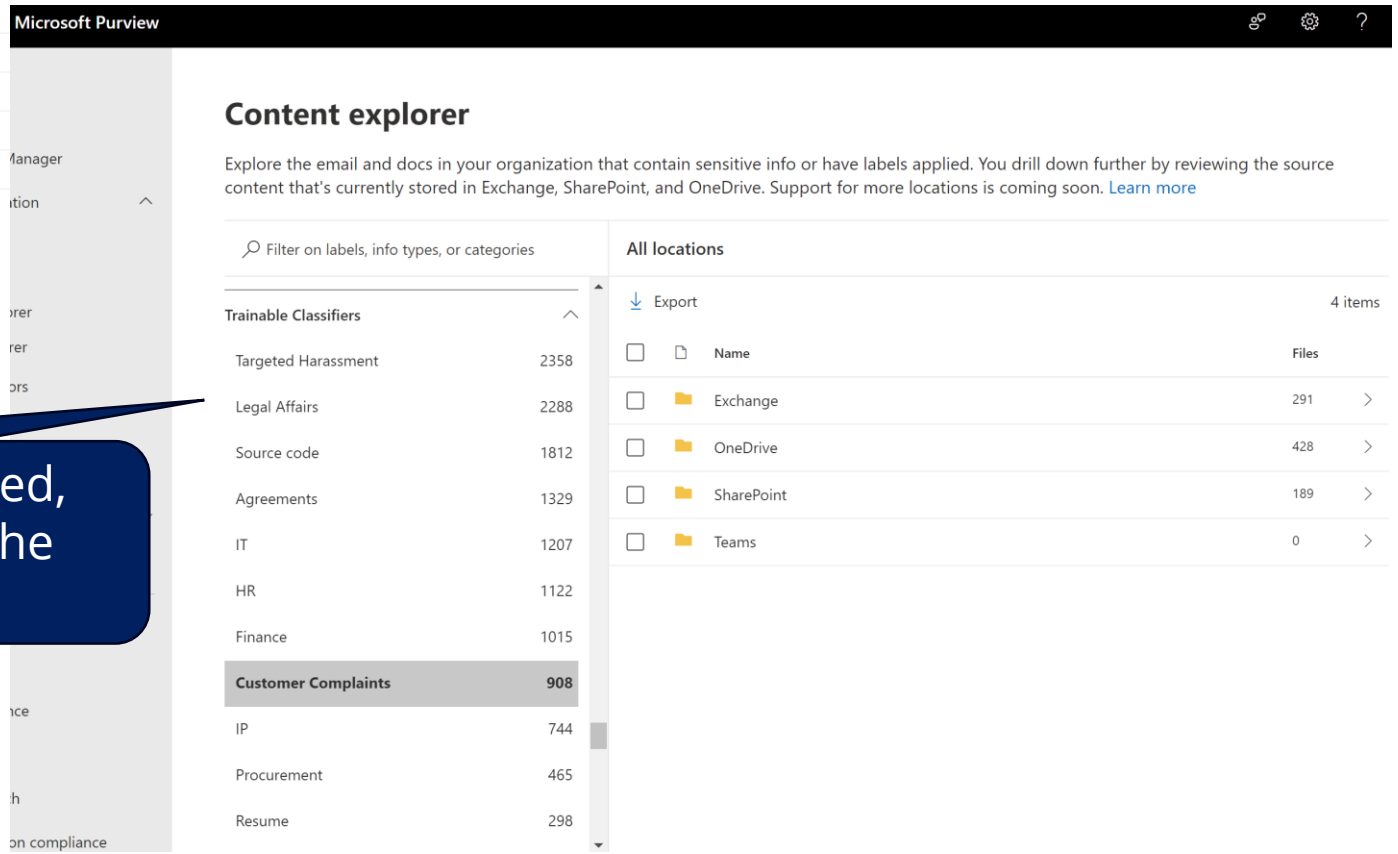
CASTLEBRIDGE

Example: Identifying Categories of Document



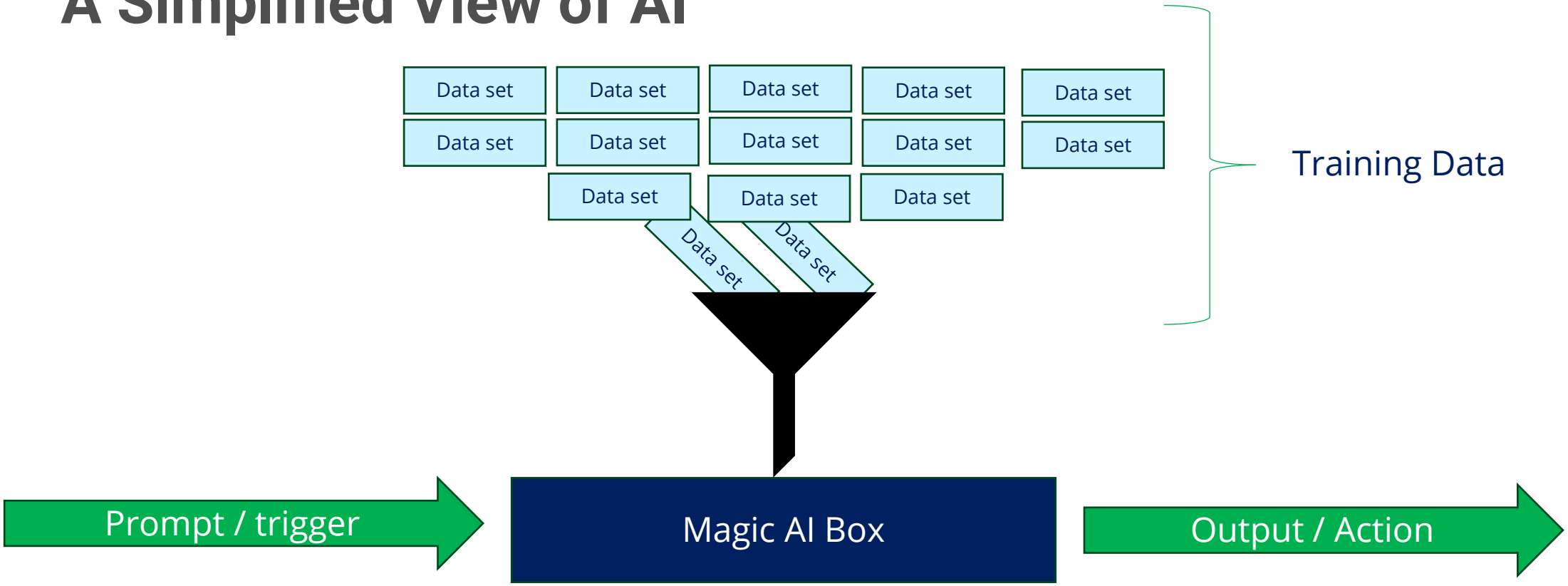
Reference patterns, Machine learning scans content to identify and tag the content as having this category of data (Unsupervised)

Supervised learning. Seed rules provided, scan content and context and *infer* if the content is one of these categories…



CASTLEBRIDGE

A Simplified View of AI



CASTLEBRIDGE

NB: AI is not magic. It is not intelligent.
It uses statistical processes to infer the next thing from all the possible next things.
It should be adopted with care, particularly in public sector contexts

LLMs> Not words or concepts but TOKENS AND PROBABILITY

Peter Piper picked a peck of pickled peppers but the peppers that Peter Piper picked were bitter, so Peter Piper had to pick a peck of better peppers

Peter Piper picked a peck of pickled peppers but the peppers that Peter Piper picked were bitter, so Peter Piper had to pick a peck of better peppers

	Pet	er	Pip	er	pick	ed	a	peck	of	pick	led	pep	per	s
Pet	1	0.8	0.2	0.8	0.2	0.5	0.65	0.2	0.35	0.2	0.1	0.1	0.1	0.85
er	0.8	1	0.8	0.1	0.8	0.8	0.1	0.65	0.1	0.1	0.2	0.3	0.3	0.625
Pip	0.2	0.8	1	0.8	0.2	0.3	0.3	0.1	0.2	0.2	0.5	0.6	0.8	0.825

Risks to Consider

- Unauthorised disclosure of data to staff using AI tool
  - Personal Data
  - Personal data of 3rd parties (not staff / customers)
  - Commercially sensitive
- Potential impact of staff personal email / documents being ingested into internal training data sets
  - Data access is restricted based on user access rights – often these are over-loose in organisations
  - User might not find things but the bot searches for patterns etc.
- Label Persistence
  - AI tool may not inherit labels/classifications applied to documents/files
  - User might not be aware of “Top Secret” classification if label is stripped!
  - AI generated content will still need to be tagged/labelled
    - POTENTIAL ENSHITTENING OF KNOWLEDGE RISK IF THIS IS AUTOMATED BADLY!!!
- Content cannot be guaranteed accurate or 100% factual
  - HUMANS are still needed!!
  - “Hallucination” still possible!
  - Potential automated “cut and paste” errors if data of other data subjects is included in auto generated content

Caveat Emptor!

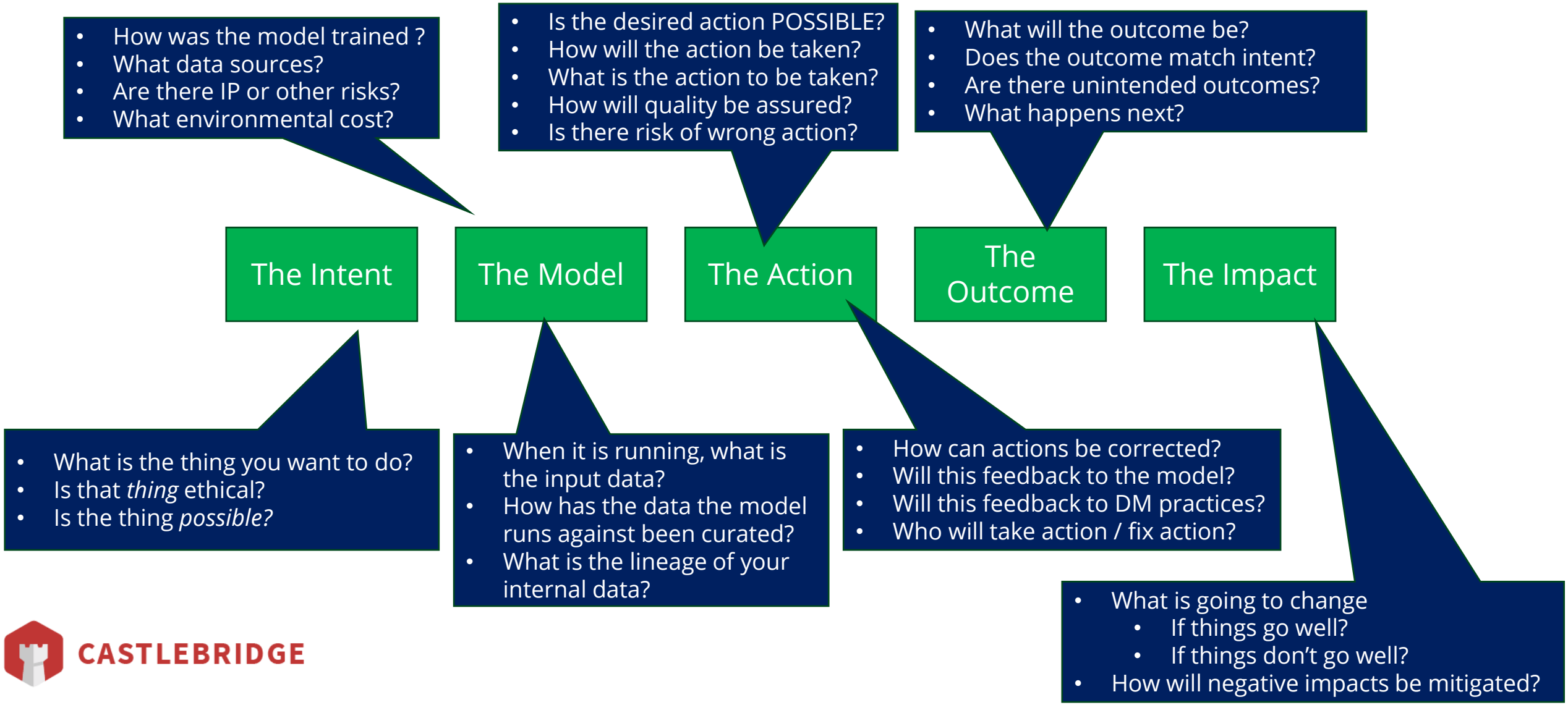
"Microsoft 365 Copilot uses your existing permissions and policies to deliver the most relevant information, building on top of our existing commitments to data security and data privacy in the enterprise. ***This means it is important to have good content management practices in the first place. For many organizations, content oversharing, and data governance can be a challenge.****"*

How to Prepare for Microsoft 365 CoPilot

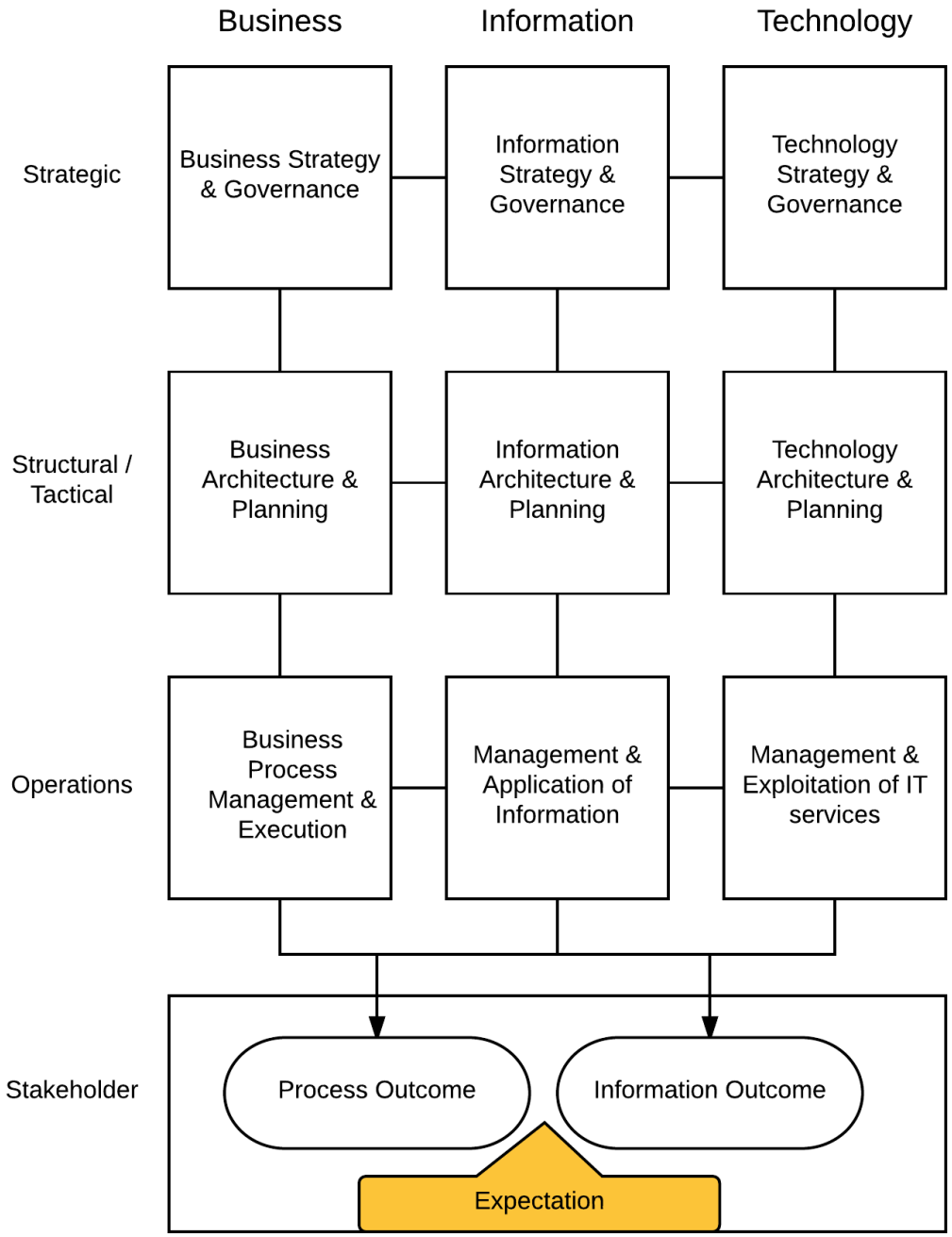

Pay no attention to that
man behind the curtain!

Ethics and AI

Ethical Life Stage Events in AI Initiatives

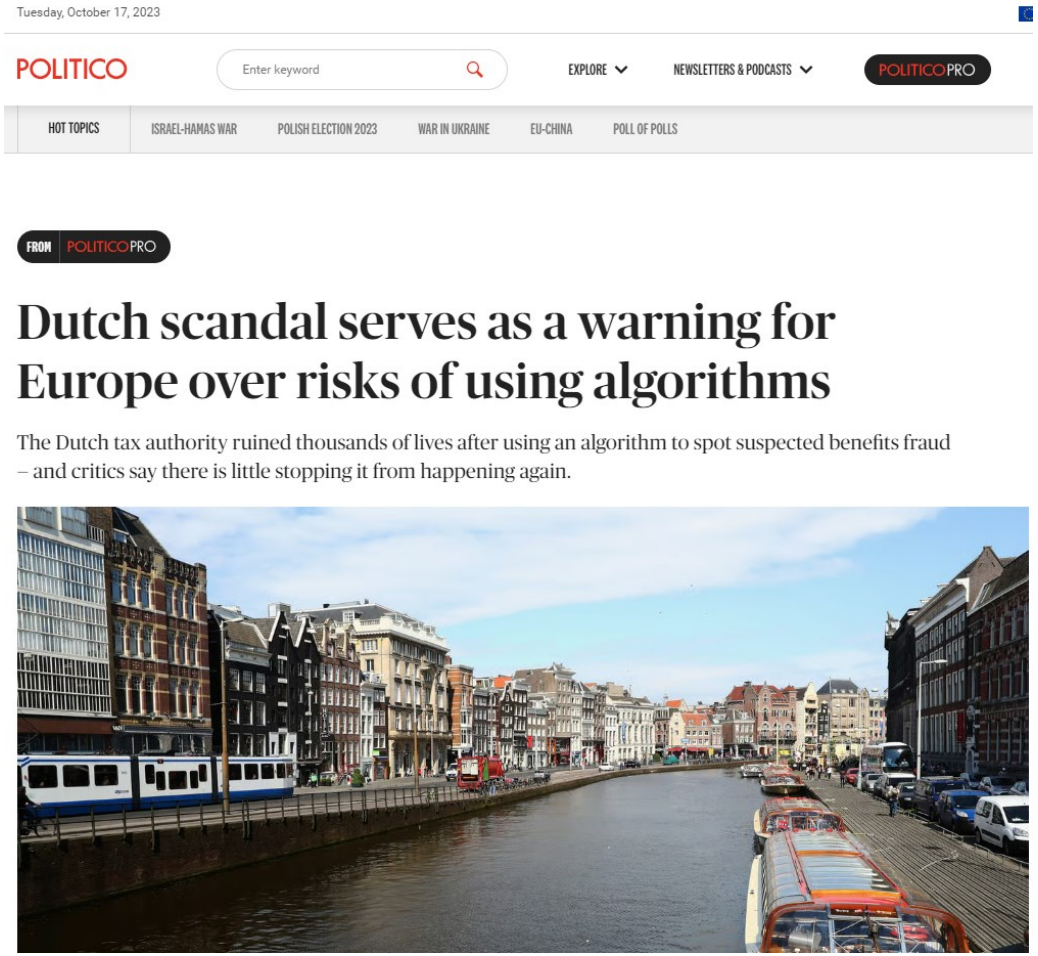


CASTLEBRIDGE



CASTLEBRIDGE

Serious Implications in Real World Applications



Tuesday, October 17, 2023

POLITICO Enter keyword EXPLORE NEWSLETTERS & PODCASTS POLITICO PRO

HOT TOPICS ISRAEL-HAMAS WAR POLISH ELECTION 2023 WAR IN UKRAINE EU-CHINA POLL OF POLLS

FROM POLITICO PRO

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

- Families penalised based on AI-derived risk indicators
- Families received tax demands that lead to poverty
- Some committed suicide
- **1000** children taken into care
- Dutch DPA fined government €3.7 million
  - Multiple violations of GDPR
  - No legal basis for processing data using AI

CASTLEBRIDGE

AI + ML

93

Air Canada must pay damages after chatbot lies to grieving passenger about discount

Airline tried arguing virtual assistant was solely responsible for its own actions

Katyanna Quach

Thu 15 Feb 2024 // 21:50 UTC

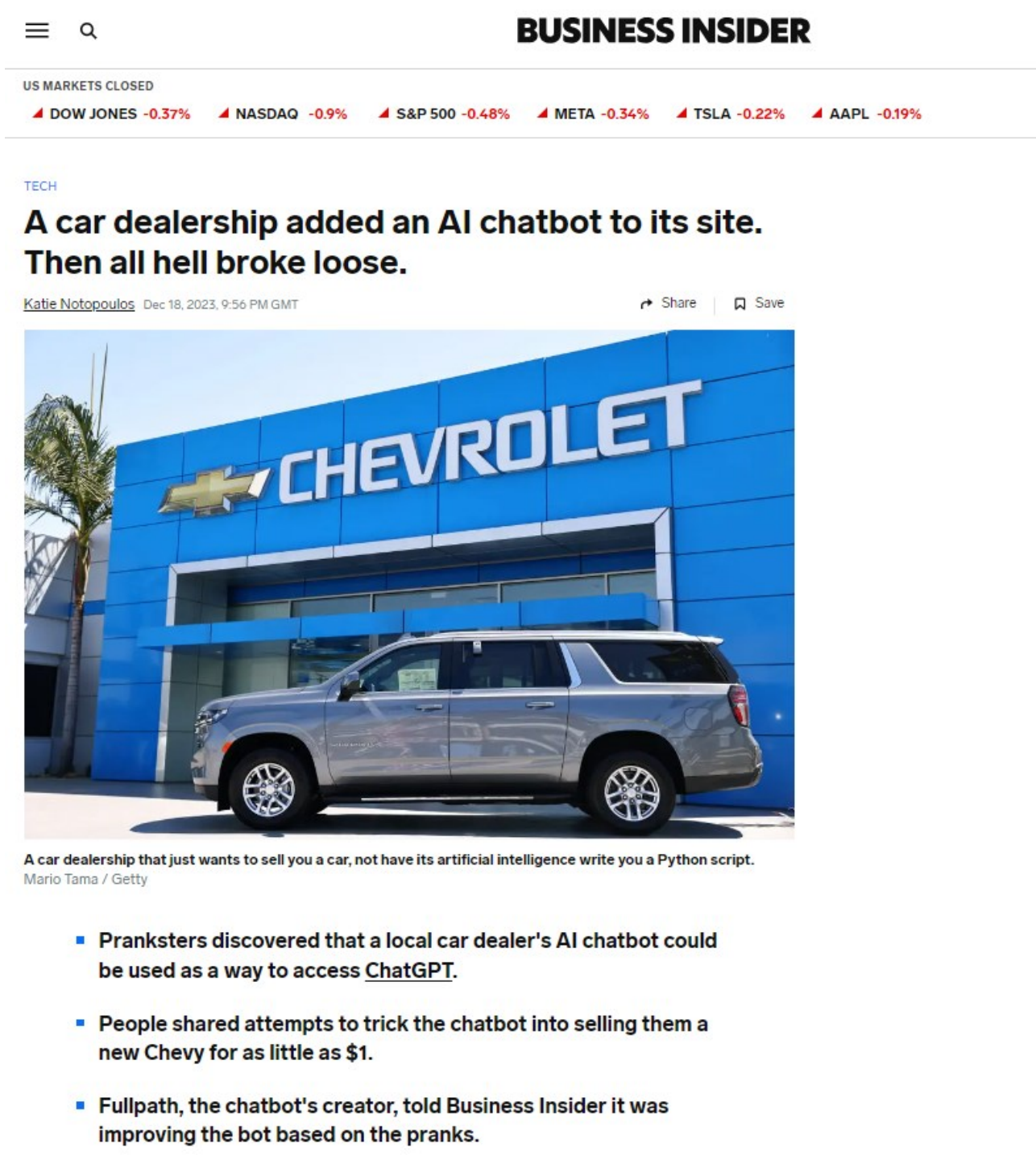
Air Canada must pay a passenger hundreds of dollars in damages after its online chatbot gave the guy wrong information before he booked a flight.

Jake Moffatt took the airline to a small-claims tribunal after the biz refused to refund him for flights he booked from Vancouver to Toronto following the death of his grandmother in November last year. Before he bought the tickets, he researched Air Canada's bereavement fares – special low rates for those traveling due to the loss of an immediate family member – by querying its website chatbot.

The virtual assistant told him that if he purchased a normal-price ticket he would have up to 90 days to claim back a bereavement discount. Following that advice, Moffatt booked a one-way CA$794.98 ticket to Toronto, presumably to attend the funeral or be with family, and later an CA$845.38 flight back to Vancouver.

I'll buy that for a dollar!

- What can we learn from Chevy and Air Canada?
- What fundamentals of data management (and IT management) should be applied to make AI implementation
  - ETHICAL?
  - EFFECTIVE?



BUSINESS INSIDER

US MARKETS CLOSED

DOW JONES -0.37% NASDAQ -0.9% S&P 500 -0.48% META -0.34% TSLA -0.22% AAPL -0.19%

TECH

A car dealership added an AI chatbot to its site. Then all hell broke loose.

Katie Notopoulos Dec 18, 2023, 9:56 PM GMT Share Save

A car dealership that just wants to sell you a car, not have its artificial intelligence write you a Python script.
Mario Tama / Getty

- Pranksters discovered that a local car dealer's AI chatbot could be used as a way to access ChatGPT.
- People shared attempts to trick the chatbot into selling them a new Chevy for as little as $1.
- Fullpath, the chatbot's creator, told Business Insider it was improving the bot based on the pranks.

CASTLEBRIDGE

Ethical Principles for AI?

Episteme (Principle)

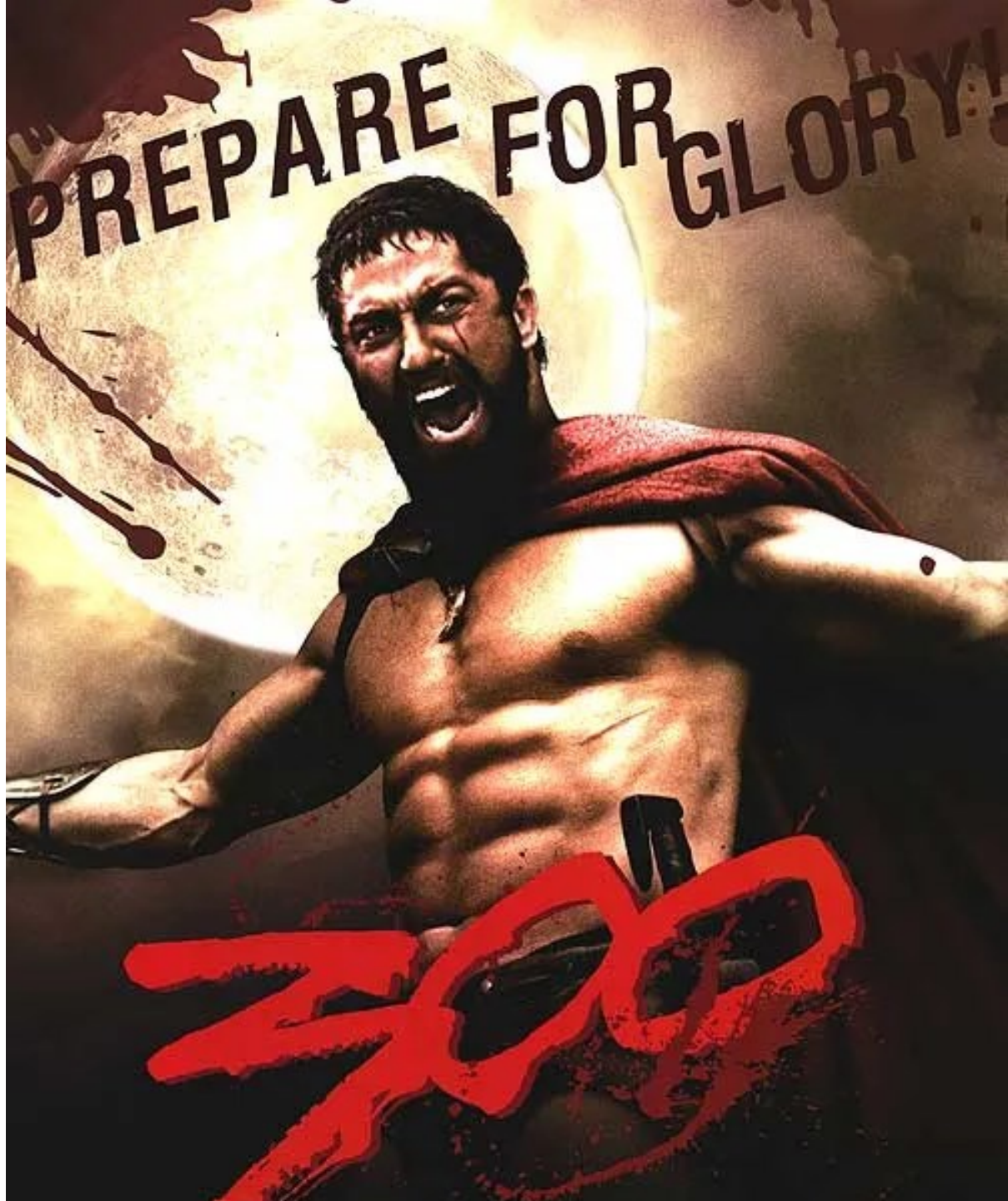
Techne (Craft)

Phronesis (Practice)

CASTLEBRIDGE

About Standards…

- Current count: 300+ standards for AI
- Deal with a range of issues
- Many are sector / application specific
- Do you know what standards apply?



CASTLEBRIDGE

Navigating Regulation

CASTLEBRIDGE

Key Regulatory Considerations

GDPR

- Transparency – Article 13 / 15
- Automated Decisions – Art 22
- Integrity of Processing– Art 32
- DP by design/default– Art 25/35
- Accountability – Art 5(2) / 24

"Special Category Data" – *OT Decision CJEU* – can now apply where inferred data.

EU AI Act

- Prohibited practices – Article 5
- "High Risk" AI – Article 6 / Annex III
- Data Governance – Article 10
- Documentation – Art. 11 / Annex IV
- Transparency – Article 13
- Transparency – Article 52

EU AI Liability Directive

Coming Soon!

CASTLEBRIDGE

Other jurisdictions are developing AI Regulatory frameworks (e.g. Ghana)

Key Risk to Consider – Platform / Vendor lock in

The cost of and disruption of changing a supplier of an AI / ML tool will be significant

What happens if your supplier goes out of business?

What happens if your support supplier goes out of business?

What happens your supplier starts using your data for training?

What happens if embedded processes are not functioning correctly?

How do you comply with Article 32 of GDPR if you cannot recover the processing capability?

CASTL

DPIAs are ESSENTIAL!!

- Processing will potentially be on a large scale
  - Key test: will it be large scale processing of *personal data*
- Processing will likely involve matching or combining of data sets
- Processing will likely be innovative or apply new technologies
  - CNIL guidance is AI is not *automatically* innovative.
    - Are you using an experimentally validated process that has been tested in real world conditions?
    - Are you using experimental techniques or methods that rely on statistical approaches or where risks are unknown (or unknowable)?
- Potentially will involve processing of special category data
  - ***OT*** case in CJEU – need to be aware of the risk of *inference* of special category data by an AI / ML process and apply appropriate safeguards

- ***Need to consider RISK to data subjects***
  - ***Does the processing result in outcomes that could impact individuals?***
  - ***Does the processing result in outcomes that could be impactful on individuals?***

- ***Need to consider different risks between TRAINING and OPERATION of a system***

CASTLEBRIDGE

A Risk Based Approach to AI is Essential!

- Is your proposed use of AI in a category of “High Risk” AI?
- Even if not high risk, are you processing data relating to people?
  - Are you drawing **inferences** about people using an AI system?
  - Are you **enriching**/correcting data based on Machine Learning/AI processes?
  - Are you making **decisions** relating to people in an automated way?
- What are the risks to your organisation if the process **doesn’t work**?
  - How can you mitigate risk of **error** in generated summaries or translations?
  - How will your ‘future state’ organisation handle **outages or exceptions**?
  - How will you correct errors (and prevent the AI systems from repeating the mistake)?
- What is the impact of bias on your outcomes?
  - Data bias
  - Model Bias
  - Human cognitive bias?
- Have you considered the impact on ‘corporate memory’ and technical / process debt if you over optimise the system?

So.. What to do?

FOMO is not a Strategy

- If you build it, they may not come

- Reasons:
  - Misaligned against Business Goals
  - Poor execution of initial implementation
  - Lack of readiness
  - Exposing defects
  - Technical and Data Debt falling due

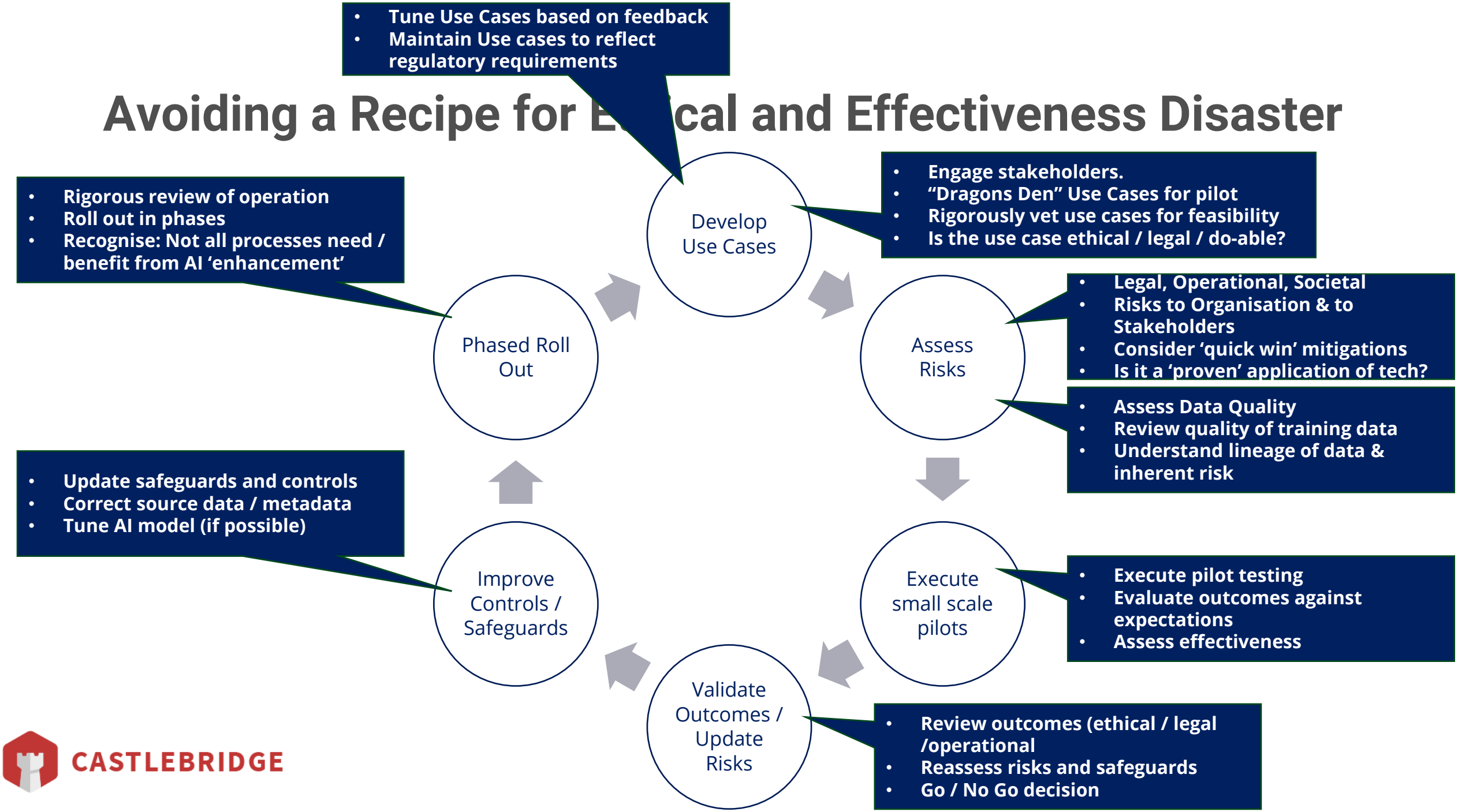
- Mitigation
  - Raise awareness of dependencies
  - Educate Stakeholders
  - Plan for exception handling / issue management

Consider your Risk Appetite (Near Term)

RISK
OF
MEDIA
ATTENTION

CASTLEBRIDGE

Avoiding a Recipe for E cal and Effectiveness Disaster

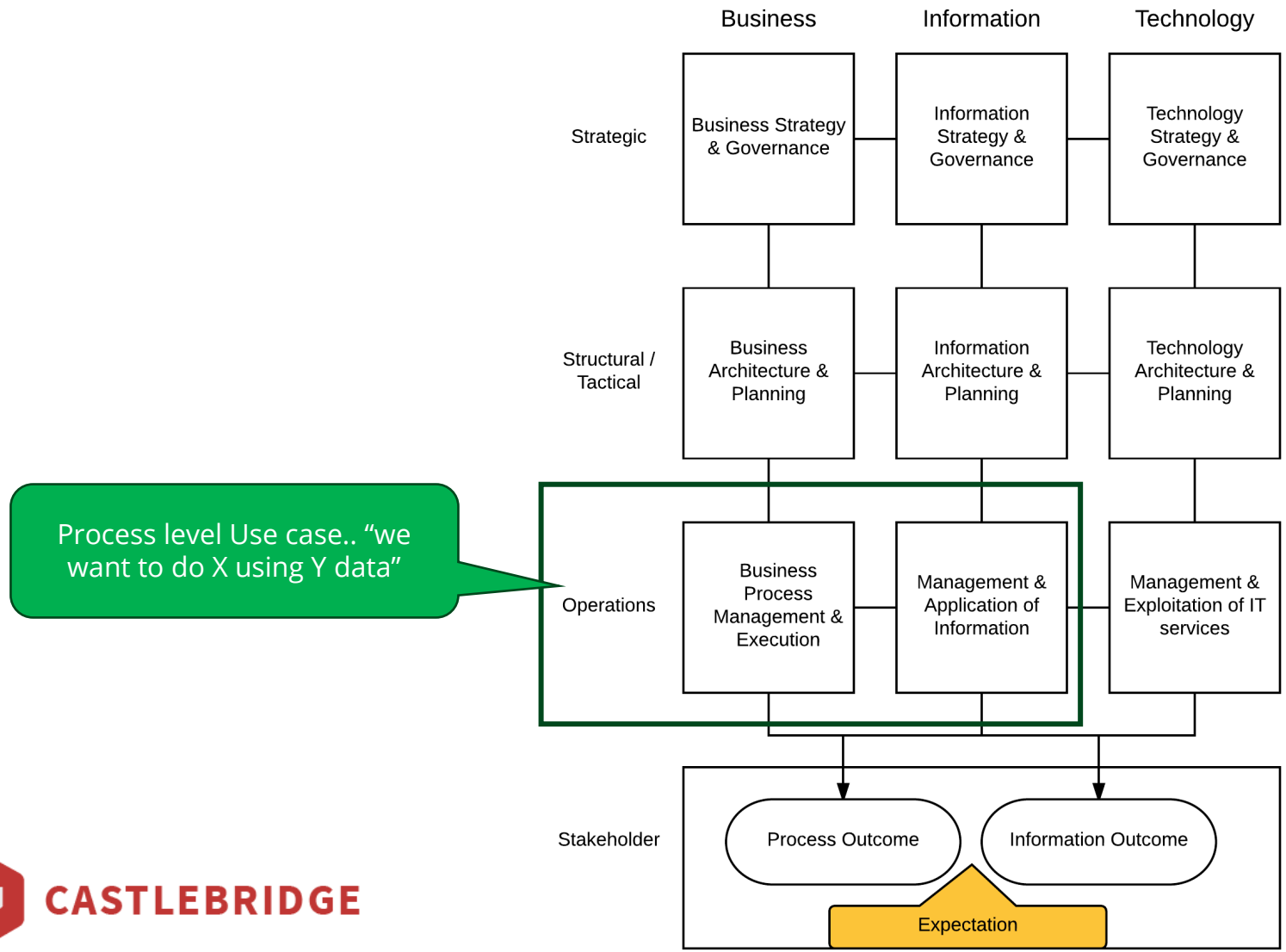


CASTLEBRIDGE

“If you don’t know where you’re going,
Any road will take you there”

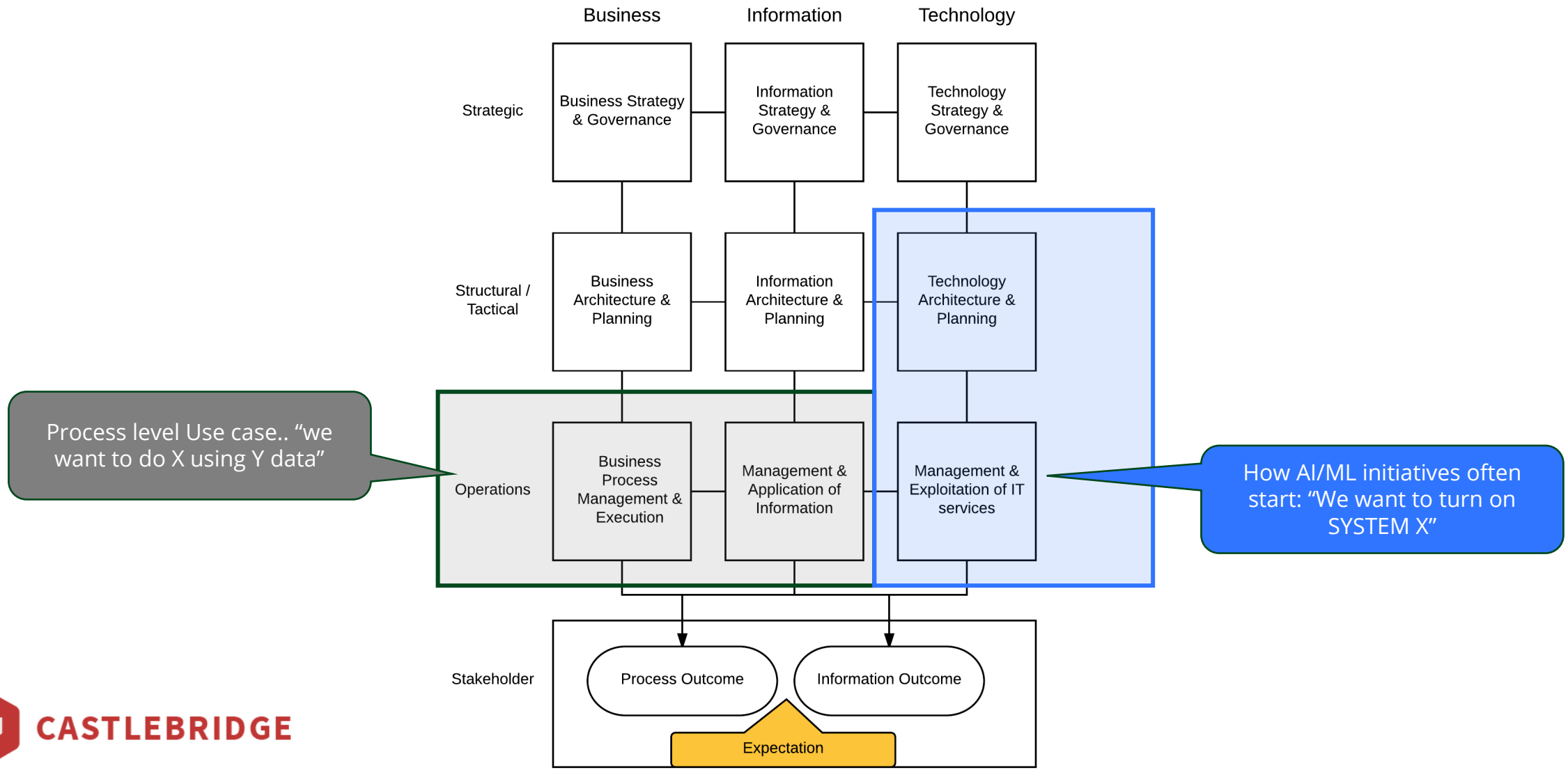
- George Harrison

Aligning for Outcomes



CASTLEBRIDGE

Aligning for Outcomes



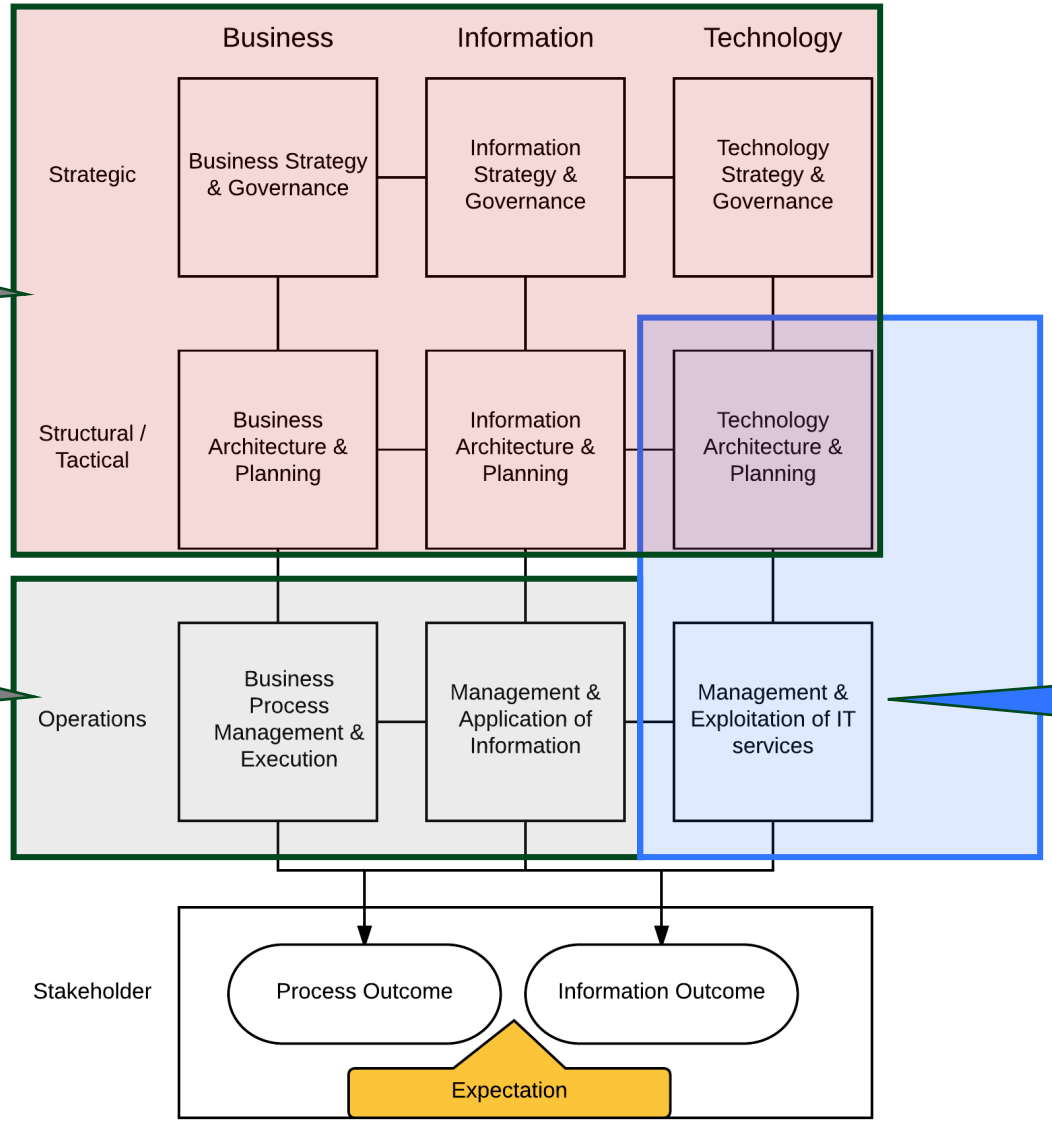
CASTLEBRIDGE

Aligning for Outcomes

Enterprise Context: Often overlooked

Process level Use case.. “we want to do X using Y data”

How AI/ML initiatives often start: “We want to turn on SYSTEM X”

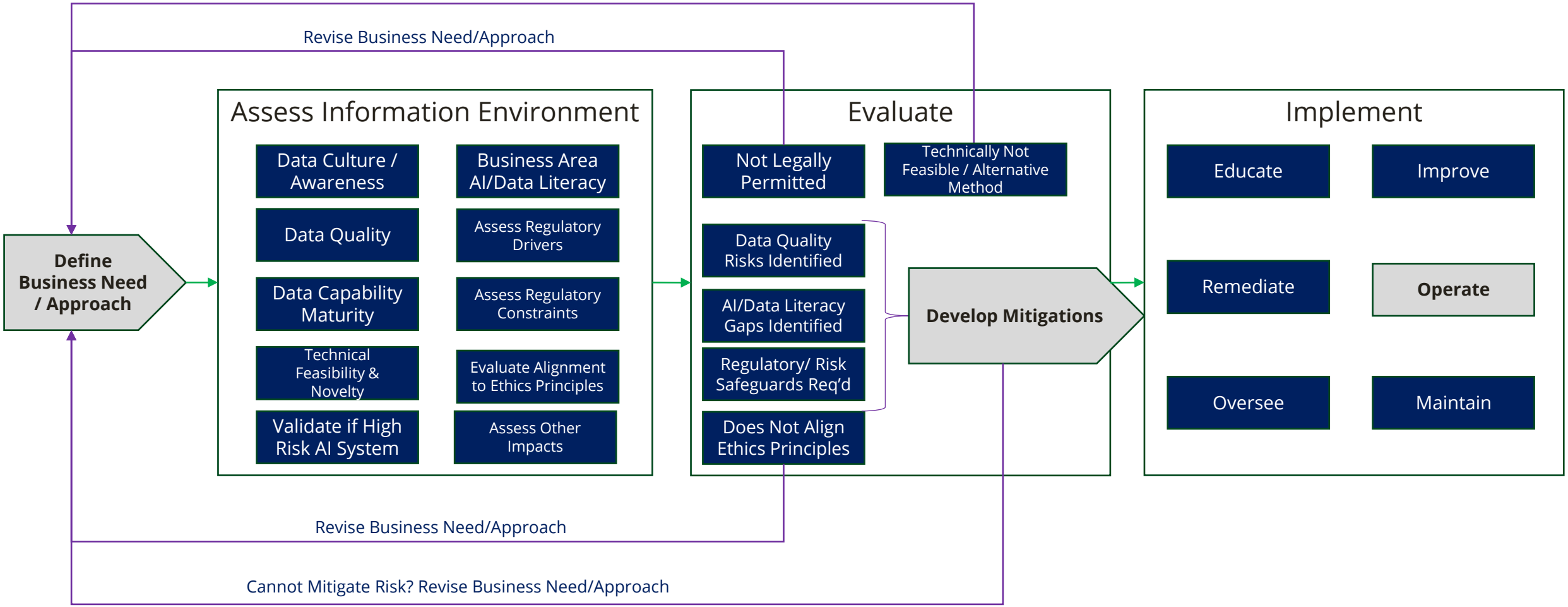


CASTLEBRIDGE

Ethical Life Stage Events in AI Initiatives

CASTLEBRIDGE

Castlebridge Impact Assessment Framework For AI

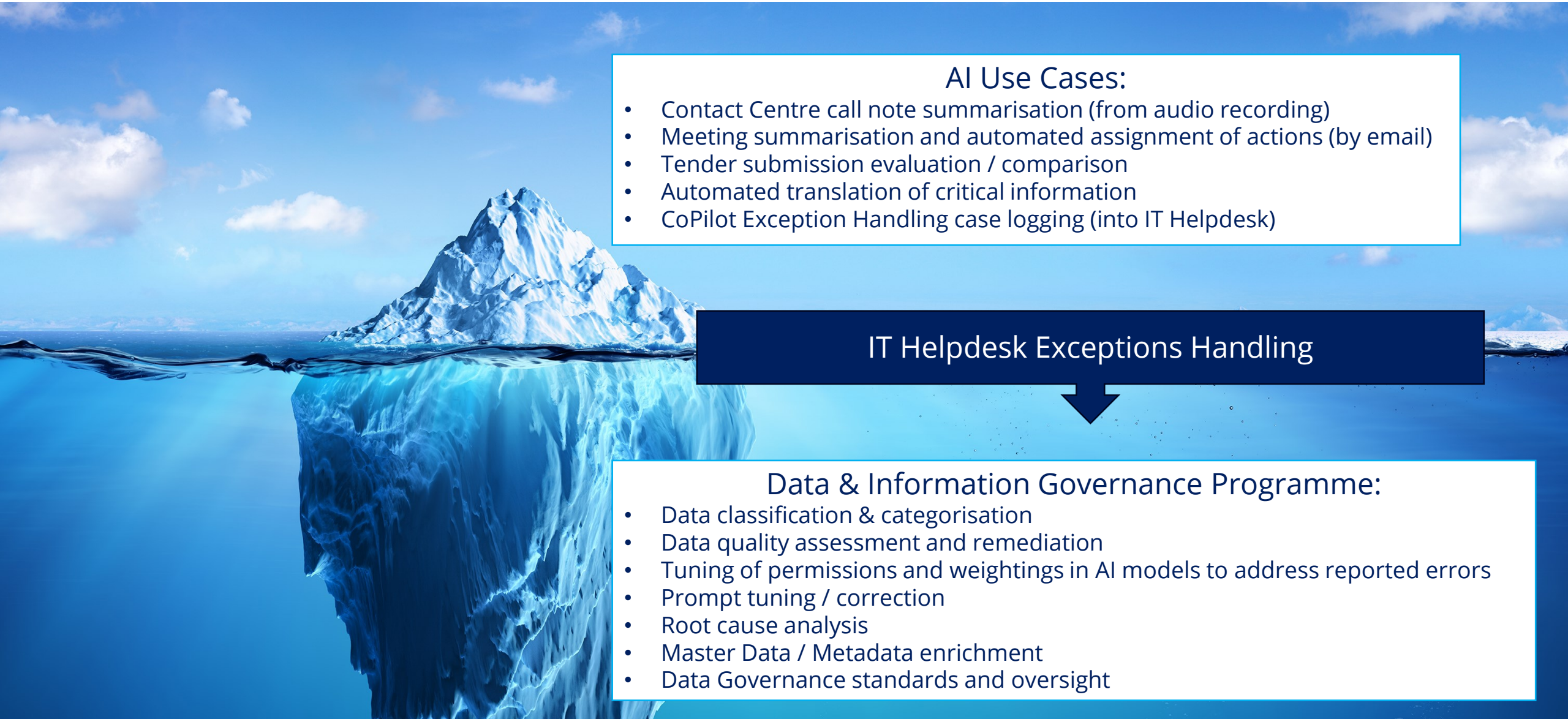


CASTLEBRIDGE

Other Simple Recommendations

1. Adopt a risk-based approach
2. Define clear **use cases** based on risk / need – be clear what is the problem/need you are addressing!
3. Be clear about the need for *ENABLING DATA DISCIPLINES* to be addressed *first as enablers*
   - Metadata management
   - Data Quality
   - Document & Content Management
   - Data Security controls /access controls
   - (Many organisations will gain significant benefits from this alone, before doing anything 'fancy')
4. Conduct rolling DPIAs/Risk Assessments
5. Learn from high profile disasters
6. **Be sceptical!**
7. Be ready to invest in humans – error correction, process design, data literacy/acumen, model tuning, quality control – all of this requires human skills!
8. Document, document, document

Example: Microsoft CoPilot in Local Government



AI Use Cases:

- Contact Centre call note summarisation (from audio recording)
- Meeting summarisation and automated assignment of actions (by email)
- Tender submission evaluation / comparison
- Automated translation of critical information
- CoPilot Exception Handling case logging (into IT Helpdesk)

IT Helpdesk Exceptions Handling

Data & Information Governance Programme:

- Data classification & categorisation
- Data quality assessment and remediation
- Tuning of permissions and weightings in AI models to address reported errors
- Prompt tuning / correction
- Root cause analysis
- Master Data / Metadata enrichment
- Data Governance standards and oversight

I love it when a plan comes together

Hannibal Smith

CASTLEBRIDGE