



The logo for Zühlke is centered on a white background. It consists of a solid purple square containing the text 'zühlke' in a white, lowercase, sans-serif font. Below 'zühlke' is the tagline 'empowering ideas' in a smaller, white, lowercase, sans-serif font. Two large, concentric dashed circles are centered behind the logo: a blue one on the left and a grey one on the right.

zühlke
empowering ideas



Lisa Falco

Consultant in Data, AI for Health & FemTech | Author of "Go Figure! The astonishing science of the female body." | Forbes contributor | Speaker on Women's Health & Medical & Responsible AI

Talks about #aiethics, #aiforgood, #womeninai, #aiinmedtech, and #aiinhealthcare

Zurich, Switzerland · [Contact info](#)

<https://www.lisafalco.com/>



Zühlke Group



EPFL (École polytechnique
fédérale de Lausanne)



Responsible AI

or: “How can we develop AI responsibly”

Introduction

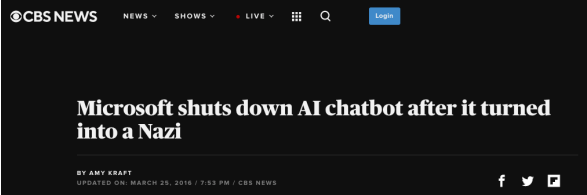


Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women.

BY JORDAN WEISSMANN OCT 10, 2018 • 4:52 PM



Amazon sign, with dude. David Ryder/Getty Images



GETTY IMAGES

Tech policy / AI Ethics

AI is sending people to jail —and getting it wrong

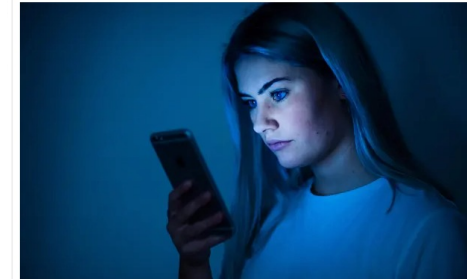
Using historical data to train risk assessment tools could mean that machines are copying the mistakes of the past.

by Karen Hao

January 21, 2019

Instagram 'pushes weight-loss messages to teenagers'

Researchers find minimal interactions by teen users can trigger a deluge of thin-body and dieting images



A teenage girl looking at her smartphone. Photograph: Alamy Stock Photo

Instagram's algorithms are pushing teenage girls who even briefly engage with fitness-related images towards a flood of weight-loss content, according to new research which aimed to recreate the experience of being a child on social networks.

Possible outcome: Poor customer experience, decreased revenues, increased cost, increased risk

These concerns are not merely about some hazy future consequences to AI bias. Organizations also cited real consequences realized from AI bias, too. A full 36% said their organization had suffered a negative impact due to an incident of AI bias in one several of their algorithms. Of those:

- 62% reported lost revenue
- 61% lost customers
- 43% lost employees
- 35% incurred legal fees dues to lawsuits or legal action, and
- 6% experienced damage to brand reputation or a media backlash

<https://www.informationweek.com/big-data/the-cost-of-ai-bias-lower-revenue-lost-customers>

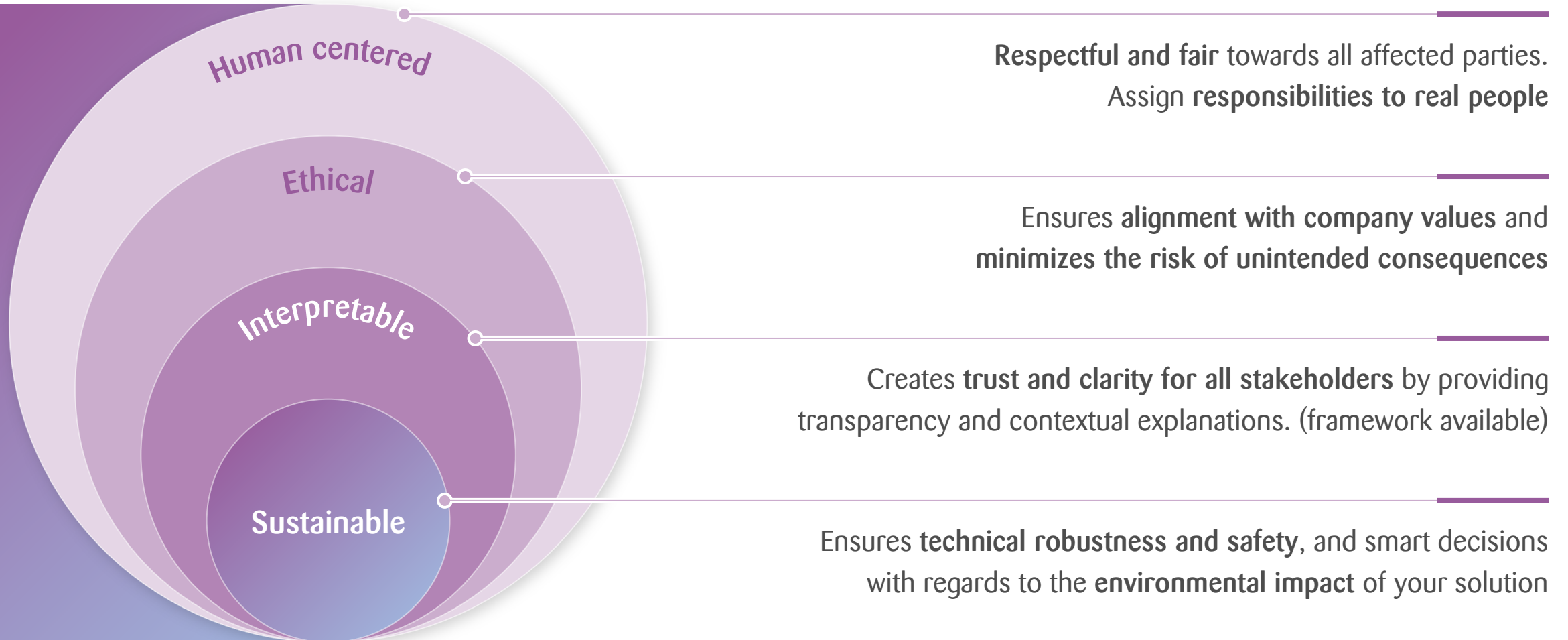
The New York Times

Here's What Happens When Your Lawyer Uses ChatGPT

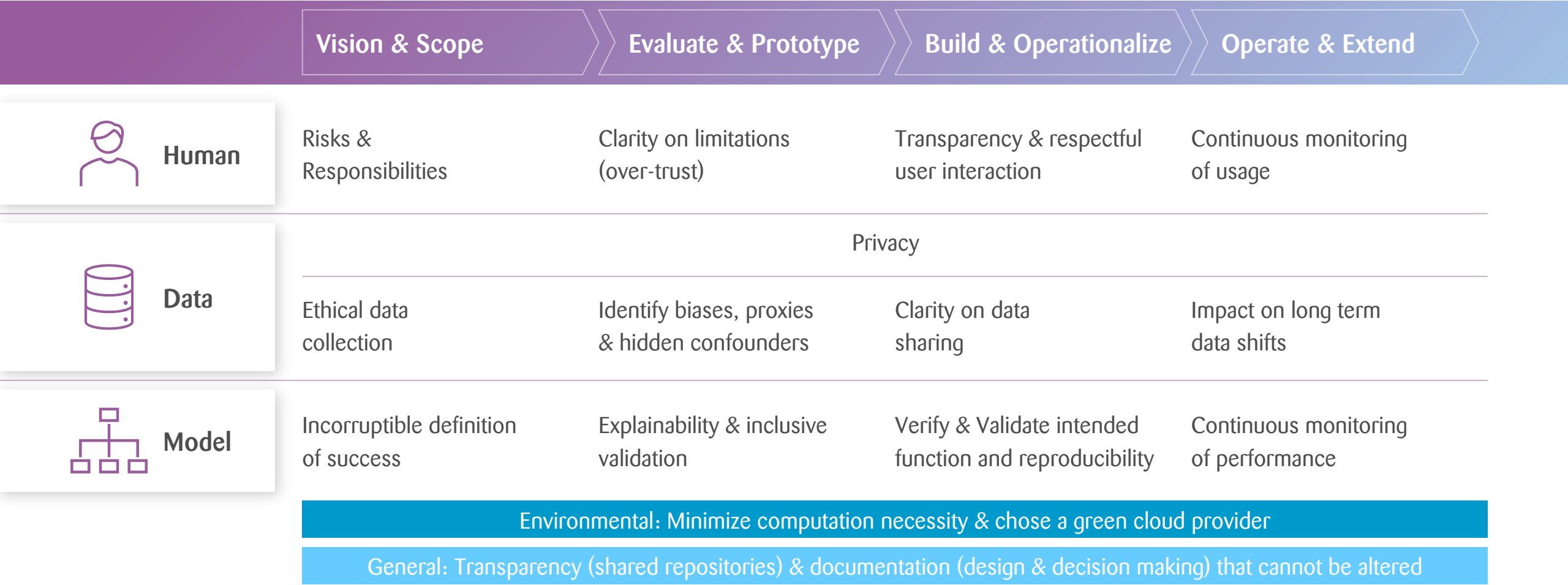
A lawyer representing a man who sued an airline relied on artificial intelligence to help prepare a court filing. It did not go well.

Core components of Responsible AI

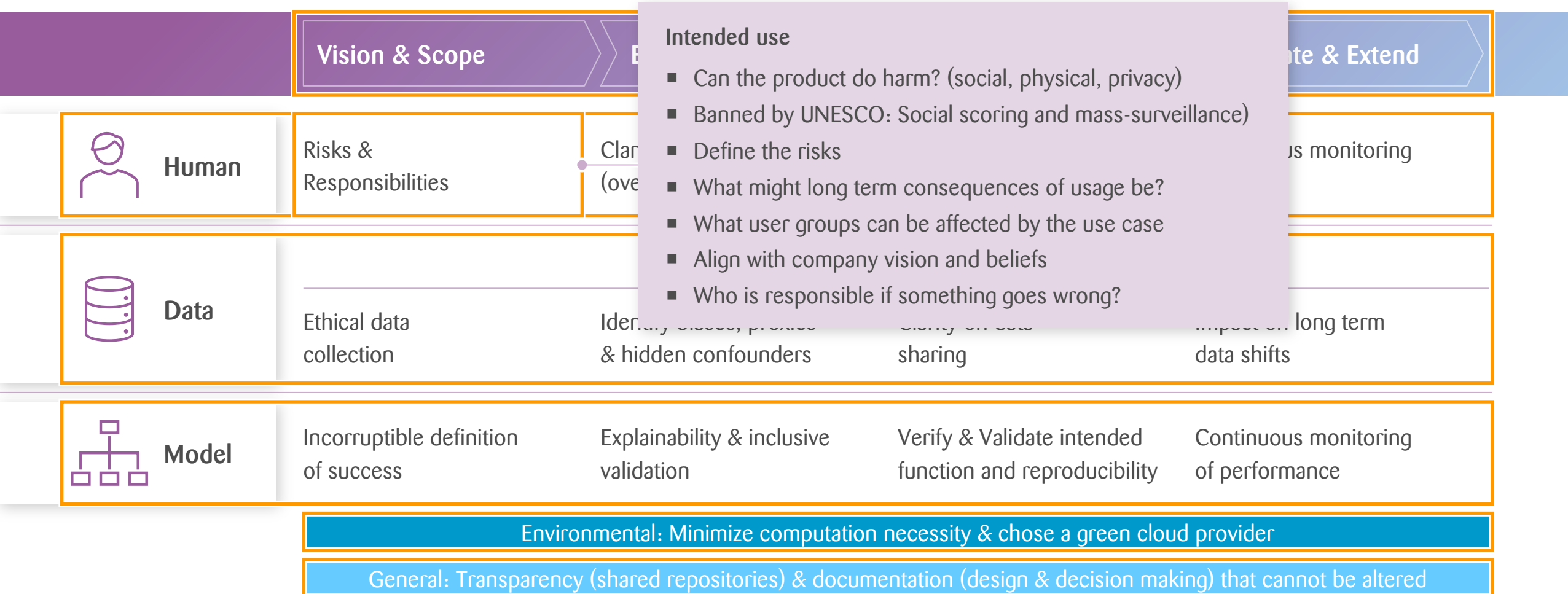
Our framework is based on a long experience of developing AI under medical regulations
& Data Governance



Zühlke's Responsible AI Framework



Zühlke's Responsible AI Framework



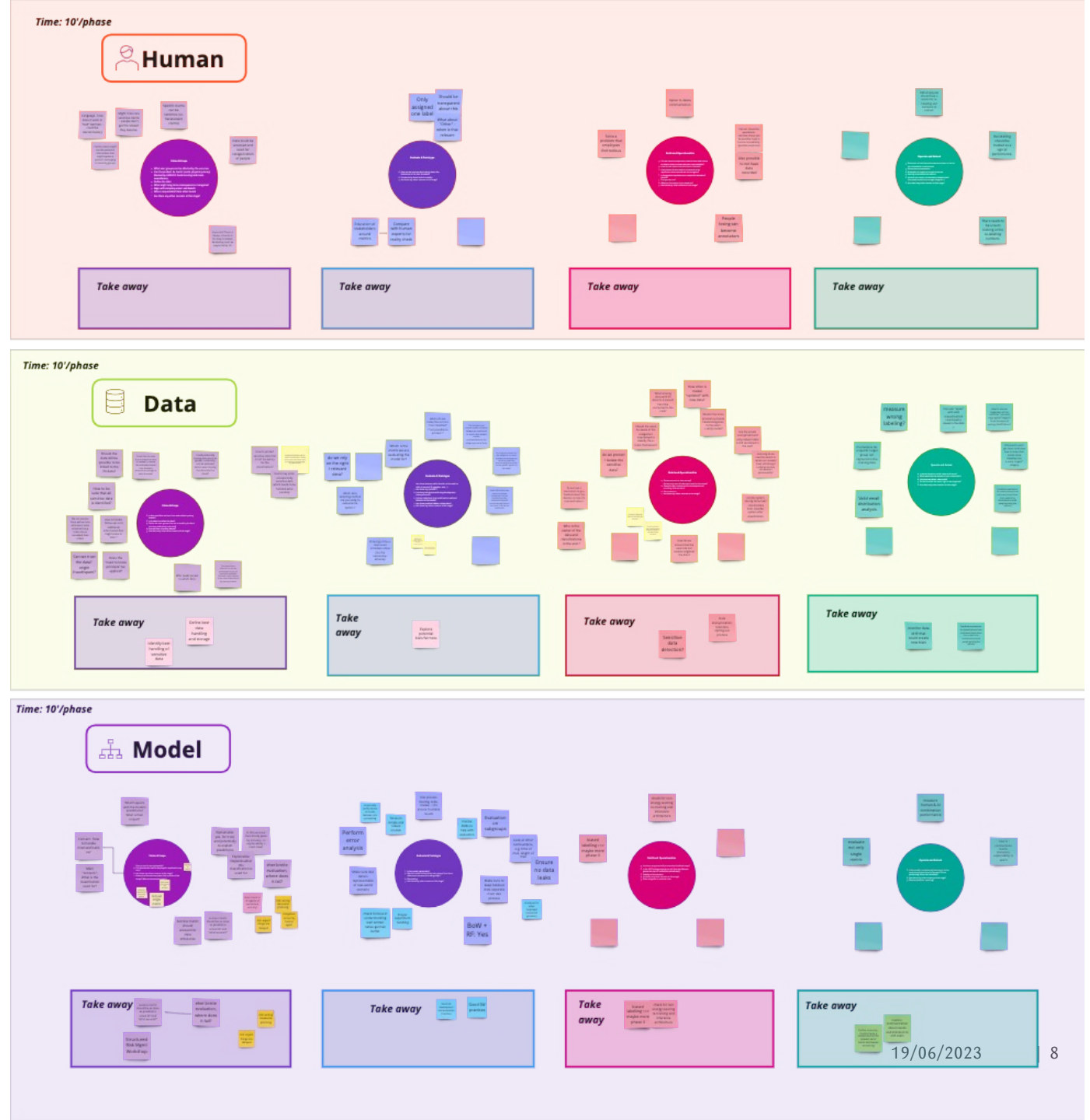
Intended use

- Can the product do harm? (social, physical, privacy)
- Banned by UNESCO: Social scoring and mass-surveillance)
- Define the risks
- What might long term consequences of usage be?
- What user groups can be affected by the use case
- Align with company vision and beliefs
- Who is responsible if something goes wrong?

Interactive workshop format

Email triage of complaints for customer support

- Conduct in groups
- Preferably diverse
- Moderation
- Favor discussion – assign enough time



Workshop outcome

Many unexpected potential issues that could lead to discrimination and bias.



Upcoming EU regulations for AI

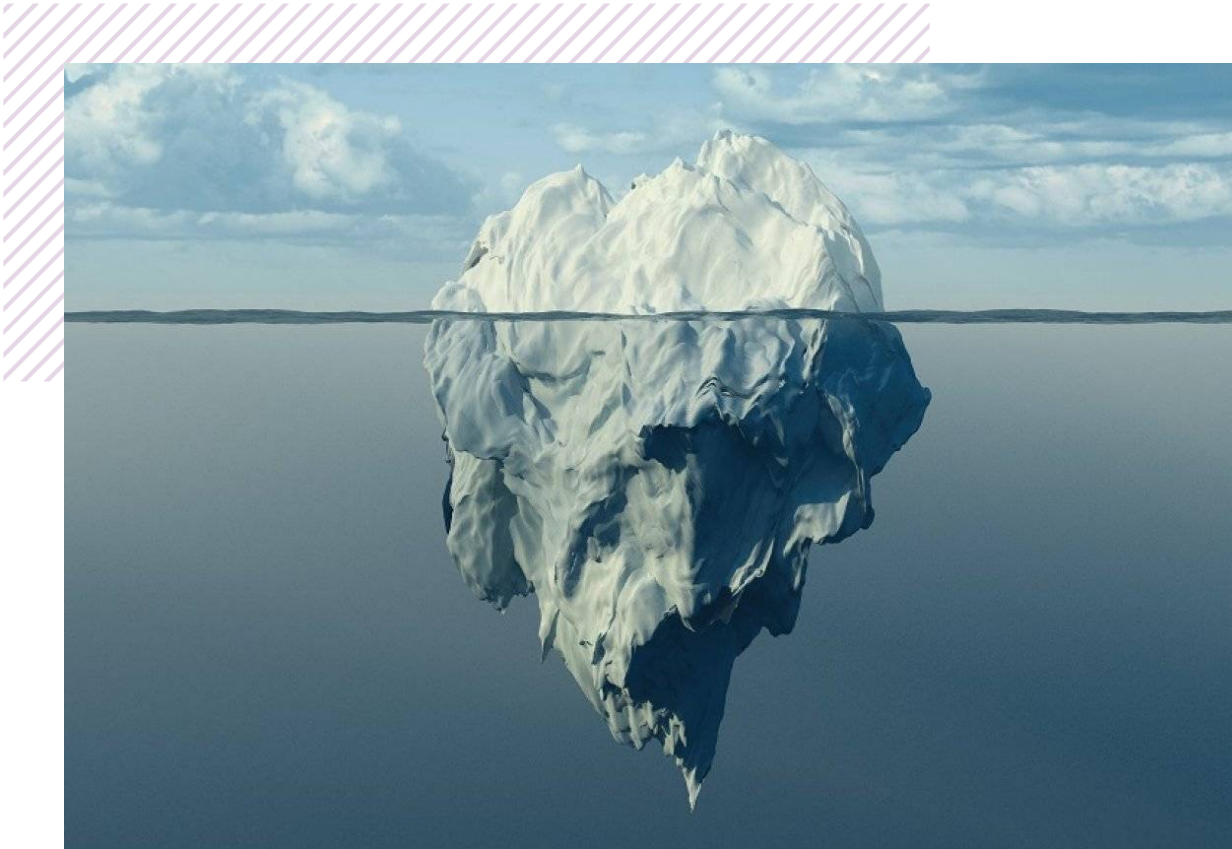
The EU AI Act is planned for end of 2023.

Risk-based and will resemble the **medical device regulations** (MDR).

Swiss companies acting in the European market will have to comply.

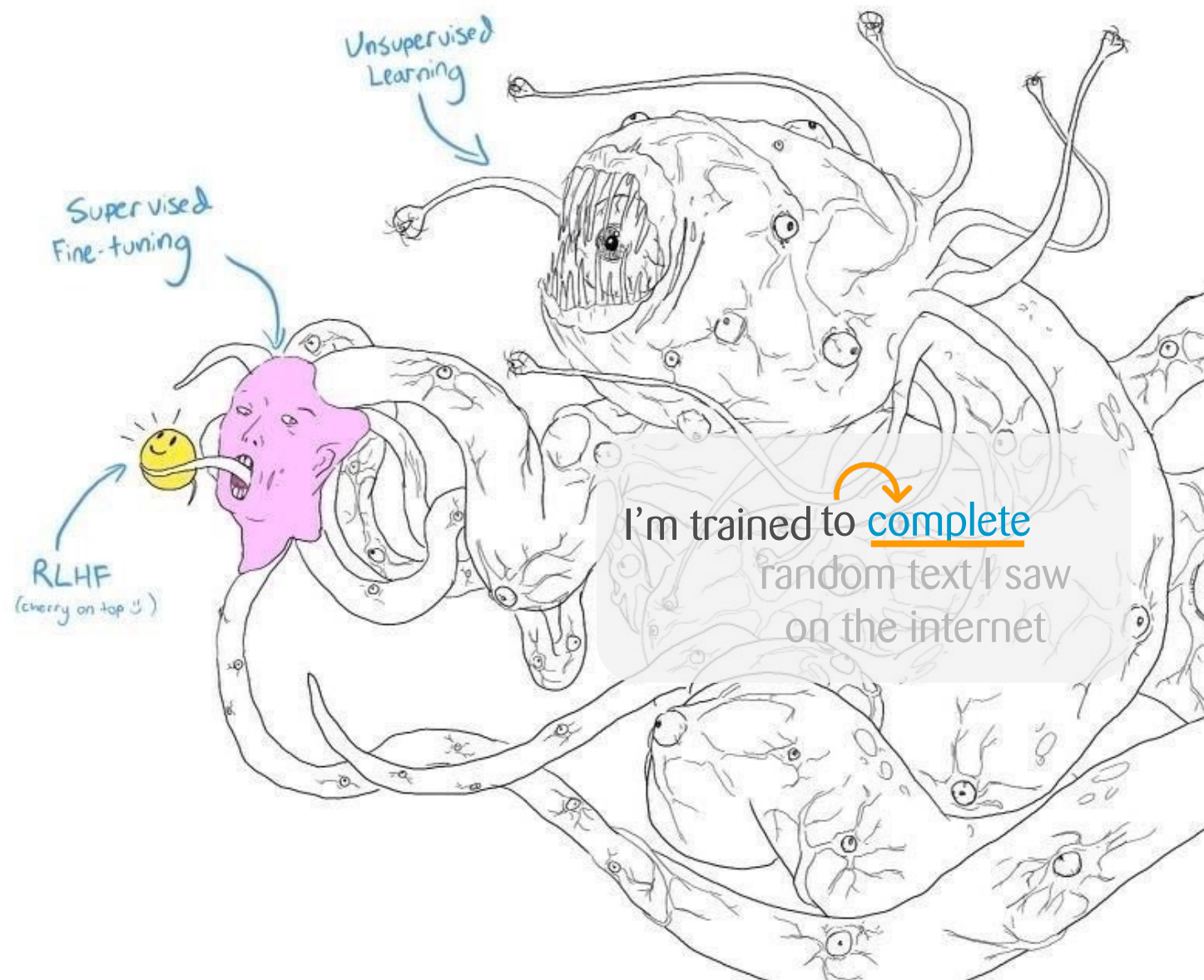


01



What about ChatGPT?

AI Alignment



An illustration of large language models from @anthrupad

Large Language Models Don't Know Things

Introducing your copilot for the web: AI-powered Bing and Microsoft Edge



\$679 million

\$3.04 billion

Discover

Chat Compose Overview Site Info

According to the web page and the web results, here is a comparison of Gap Inc. and Lululemon in the third quarter of fiscal 2022:

Metric	Gap Inc.	Lululemon
Net sales	\$4.04 billion ¹	\$1.9 billion ²
Net sales growth	2% ¹	28% ²
Gross margin	37.4% ¹	58.7% ²
Operating margin	5.9% ¹	20.7% ²
Diluted earnings per share	\$0.42 ¹	\$1.65 ²
Cash and cash equivalents	\$1.4 billion ¹	\$352.6 million ²
Inventory	\$1.9 billion ¹	\$1.7 billion ²

Learn more: [click to view](#)

<https://dkb.blog/p/bing-ai-cant-be-trusted>

Gross margin does not appear in the document. Fully hallucinated. Real value is 55.9%

Operating margin is 19%

\$2.00 not \$1.65

Large Language Models Don't Know Things

Introducing you
powered Bing and



FORBES > INNOVATION > CONSUMER TECH

Lawyer Uses ChatGPT In Federal Court And It Goes Horribly Wrong

Matt Novak Contributor ⓘ
*FOIA reporter and founder of Paleofuture.com,
writing news and opinion on every aspect of...*

Follow

May 27, 2023, 06:11pm EDT

\$3

\$3.04 billion

Discover

Chat Compose Overview Site Info

According to the web page and the web results, here is a comparison of Gap Inc.

ss margin does not appear
n the document. Fully
allucinated. Real value is
55.9%

Operating margin is 19%

\$2.00 not \$1.65

Inventory

\$1.9 billion 1	\$1.7 billion 2
--------------------	--------------------

Learn more: [click to view](#)

https://dkb.blog/

Risks of Large Language Models



False Statements With High Confidence



Biased & Harmful Generations



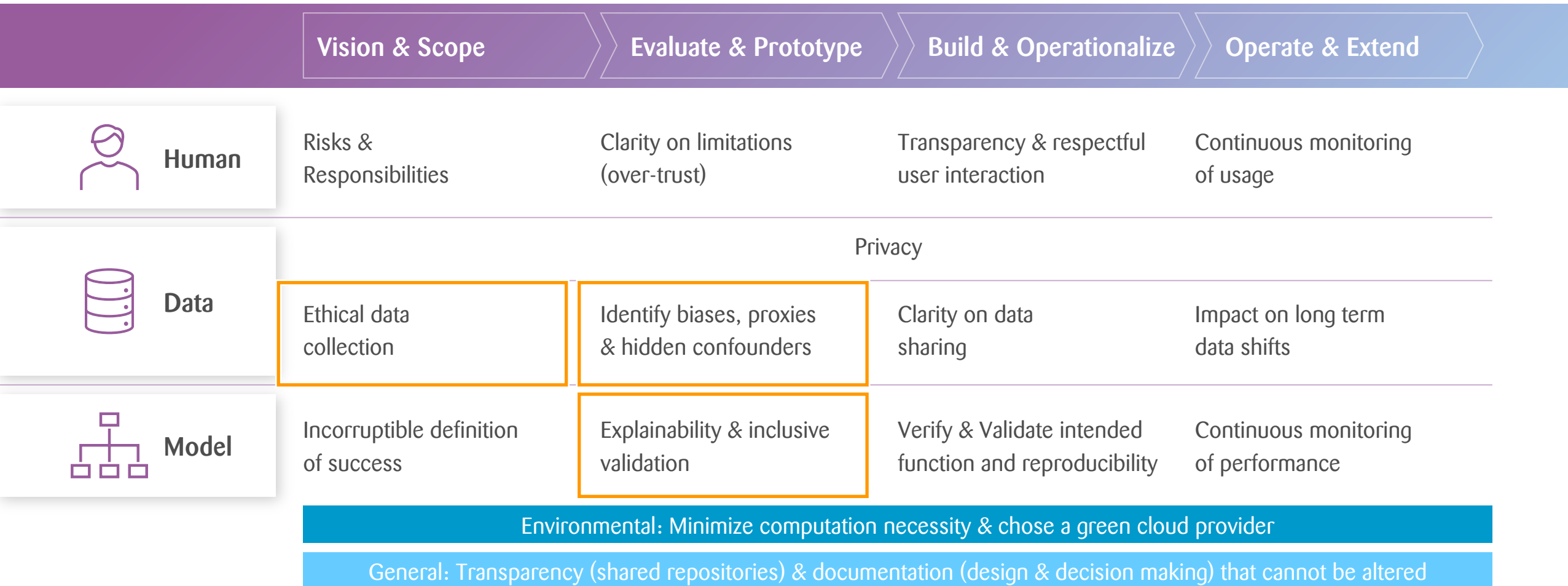
Legal Uncertainties



Misuse

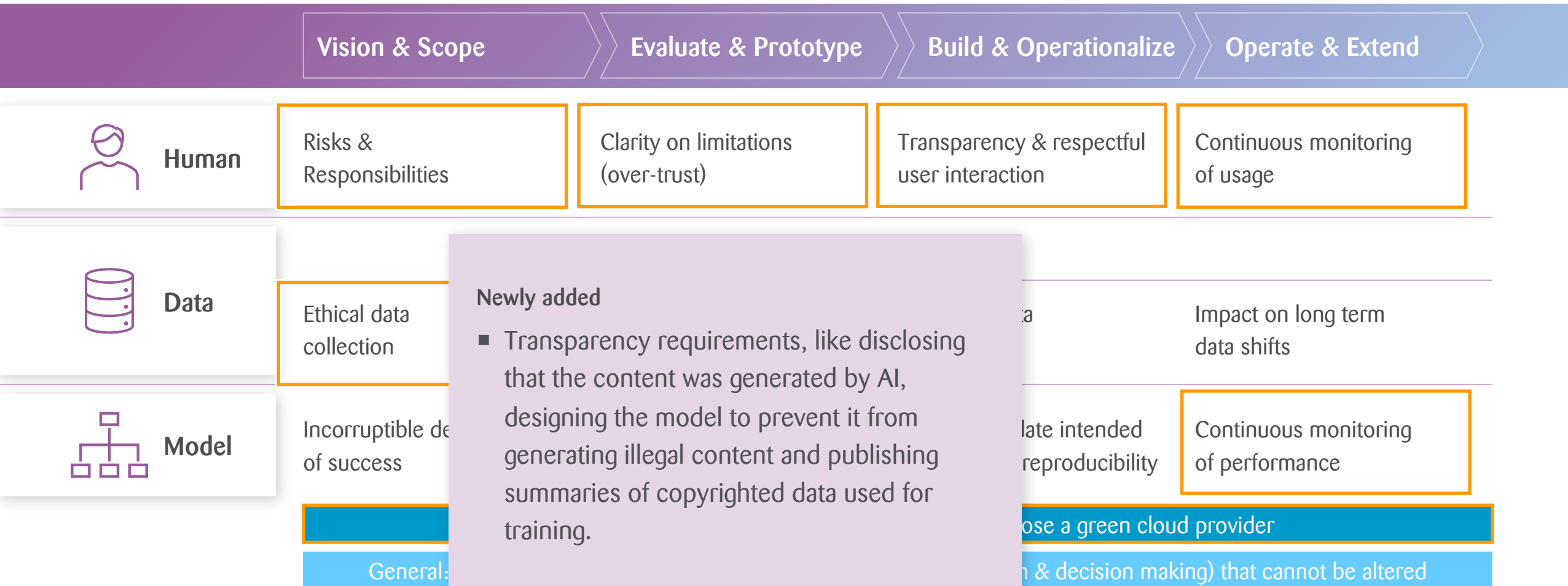
Zühlke's Responsible AI Framework – Generative AI

What can you do to avoid the pitfalls when implementing generative AI.



Zühlke's Responsible AI Framework – Generative AI

What can you do to avoid the pitfalls when implementing generative AI.



Thank you!



Lisa Falco

Lead Consultant Data & AI

Mobile +41 79 412 2659

lisa.falco@zuehlke.com

[lisafalco.com](https://www.lisafalco.com)

LinkedIn: lisa-falco-jon

Thanks!

Responsible AI

Kubernetes days 15.06.2023