AI AND THE ART OF MANIPULATION

How We Need to Think Differently About AI as We Develop Socially Responsible Applications

How Artificial Intelligence and Machine Learning
Transform the Human Condition
JULY 20, 2021

ANDREW MAYNARD
Professor and Associate Dean
Arizona State University College of Global Futures

AI ETHICS

2017

Asilomar AI Principles

Future of Life Institute

https://futureoflife.org/ai-principles/

2018

Everyday Ethics for Artificial Intelligence

IBM

https://www.ibm.com/design/ai/ethics/everyday-ethics/

2019

Ethically Aligned Design:

A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems

IEEE

https://ethicsinaction.ieee.org/

2020

AI Now
New York University
https://medium.com/@AINowInstitute/ai-in-2020-a-year-to-give-us-pause-67795fe23324

"2020 has been a year of hard truths and tragedy, as interlocking crises put the failures, inadequacies, and structural limitations of our core institutions in the spotlight. **At the same time, we see the AI industry rushing to profit in the space left by an absent social safety net, bolstered by governments' increasing turn to tech solutions.** AI companies are ramping up surveillance of our workplaces, schools and communities; cracking down on worker organizing and ethical research; and bankrolling the passage of bills that gut worker protections for millions — while growing richer and more powerful in the process."

(Emphasis added)

2021

National Artificial Intelligence Initiative

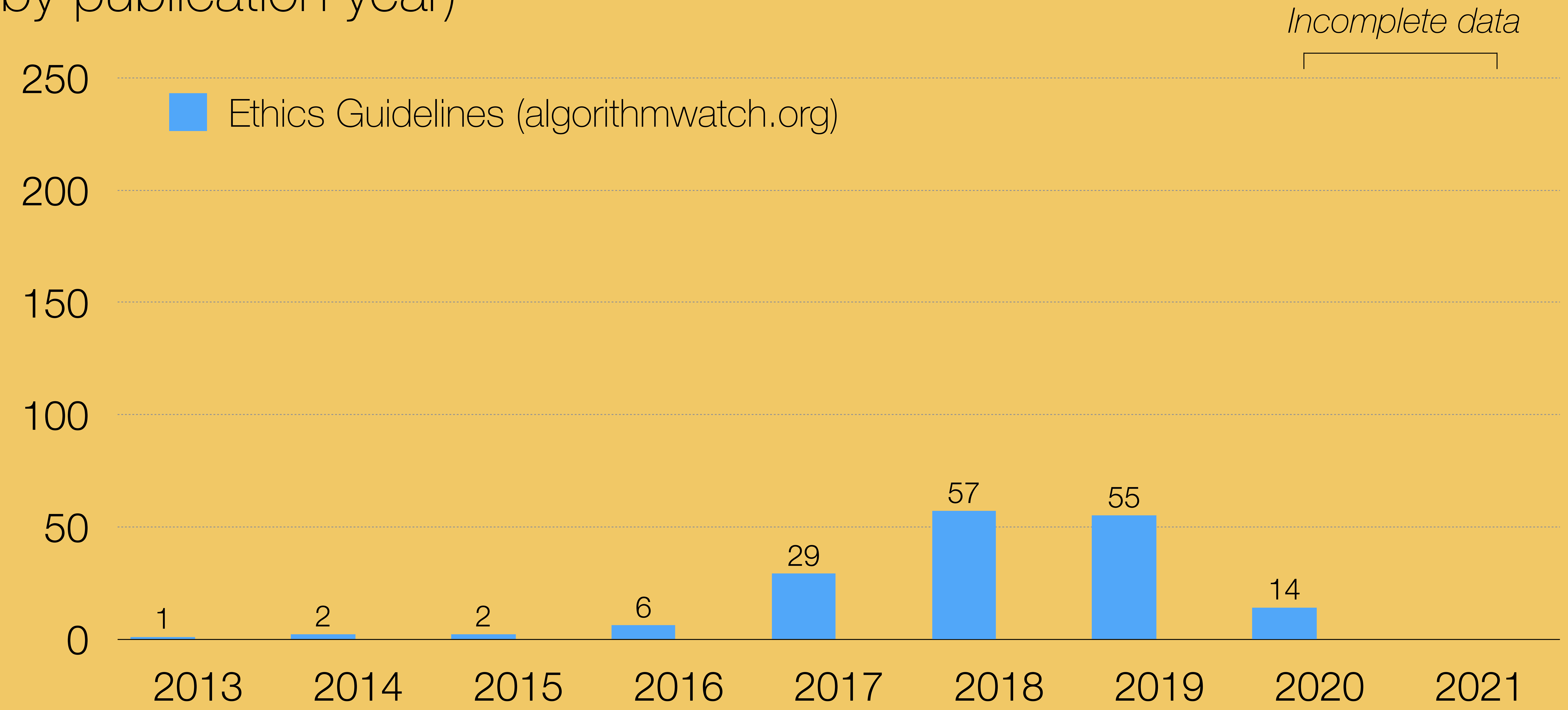
Including “Advancing Trustworthy AII”

US federal Government

https://www.ai.gov/

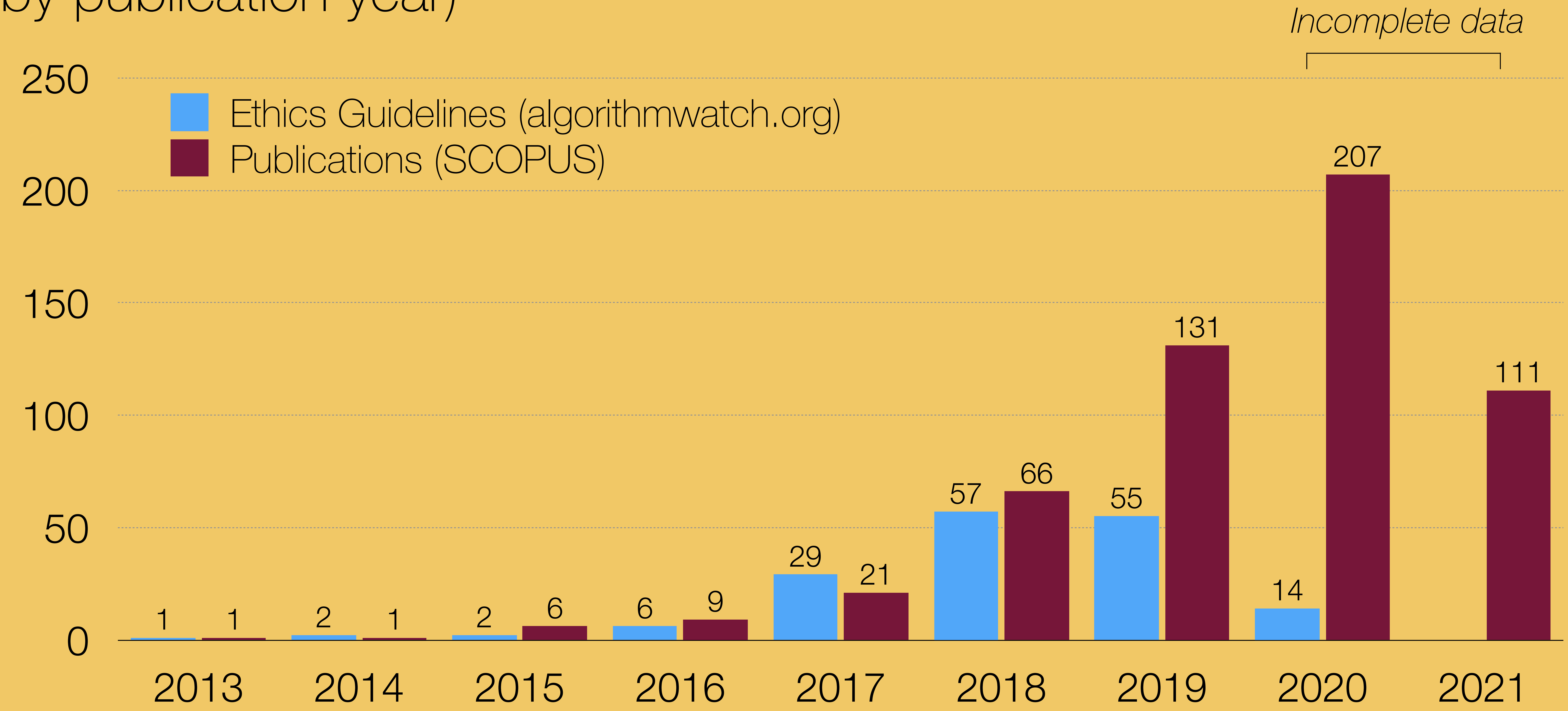
AI Ethics Guides and Publications

(by publication year)


Incomplete data

AI Ethics Guides and Publications

(by publication year)



SCOPUS search: TITLE (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning") AND (ethic*))

IEEE General Principles

Commonalities with many other sets of principles

Human Rights

A/IS* shall be created and operated to respect, promote, and protect internationally recognized human rights.

Well-being

A/IS creators shall adopt increased human well-being as a primary success criterion for development.

Data Agency

A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.

Effectiveness

A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.

Transparency

The basis of a particular A/IS decision should always be discoverable.

Accountability

A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.

Awareness of Misuse

A/IS creators shall guard against all potential misuses and risks of A/IS in operation.

Competence

A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

*Autonomous/Intelligent Systems

Source: https://ethicsinaction.ieee.org/

Are ethical frameworks enough to ensure safe and beneficial development and applications of machine learning and other forms of AI?

RISKS OF AI

Data

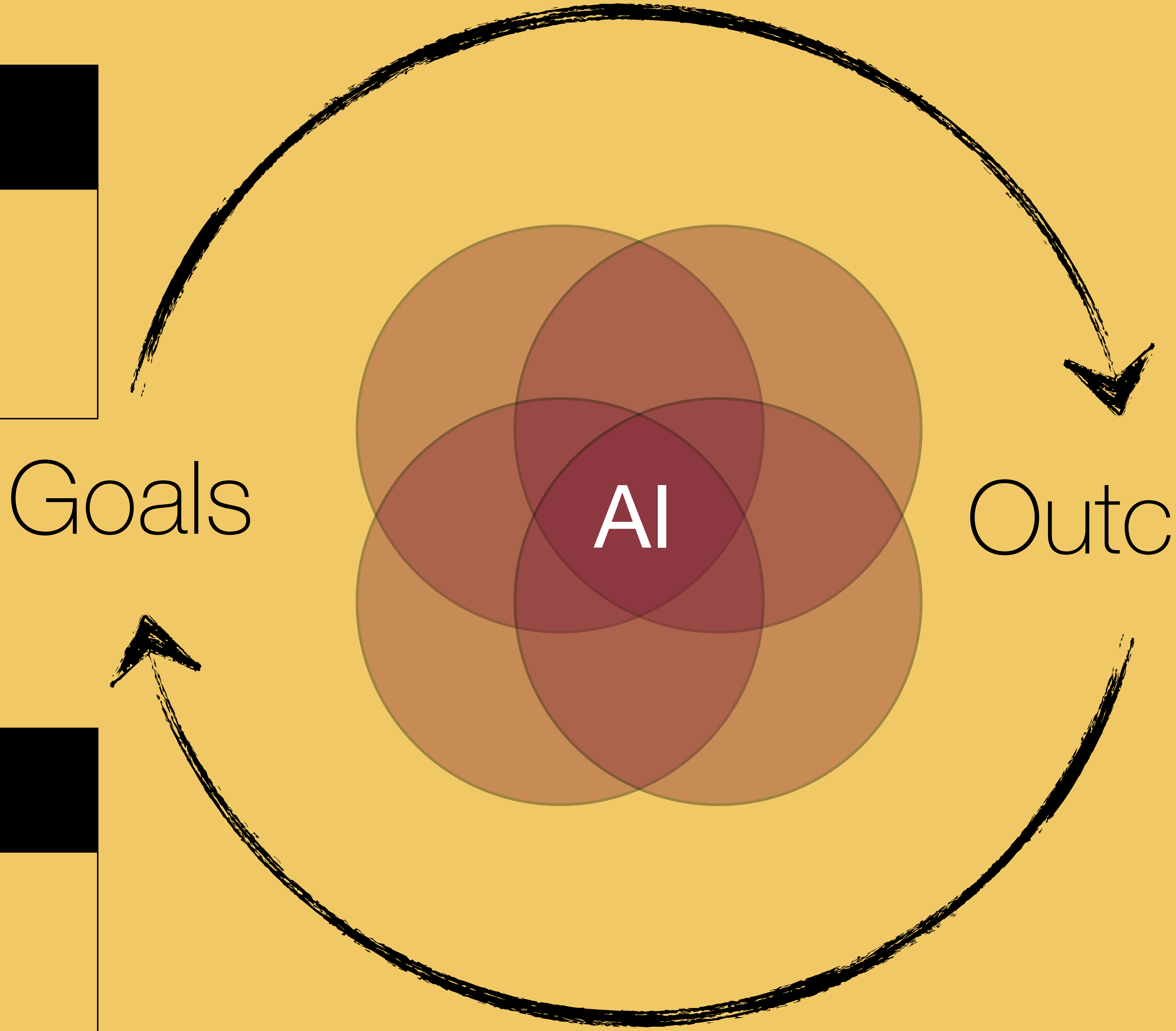
Curated
Uncurated
Etc…

Closed
Open

“Knowledge”

Deterministic
Inferential
Etc…

Intuitive
Complex



Mechanisms

Digital
Mechanical
Social

Behavioral
Political
Etc…

Constraints

Operational
Legal
Etc…

Institutional
Ethical

Data

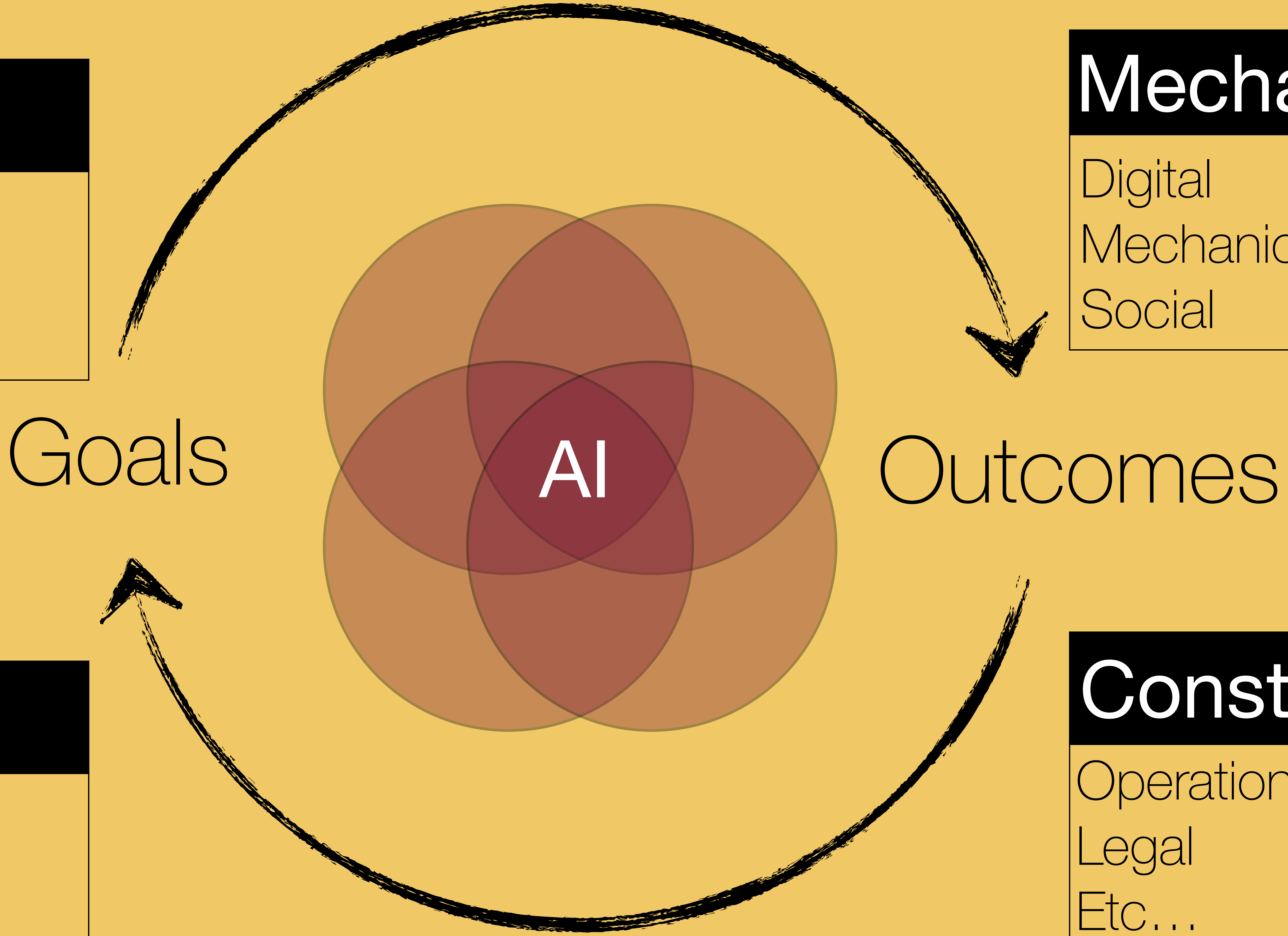
Curated
Uncurated
Etc…

Closed
Open

"Knowledge"

Deterministic
Inferential
Etc…

Intuitive
Complex



Mechanisms

Digital
Mechanical
Social

Behavioral
Political
Etc…

Constraints

Operational
Legal
Etc…

Institutional
Ethical

Wrong ← **ETHICS** → Right

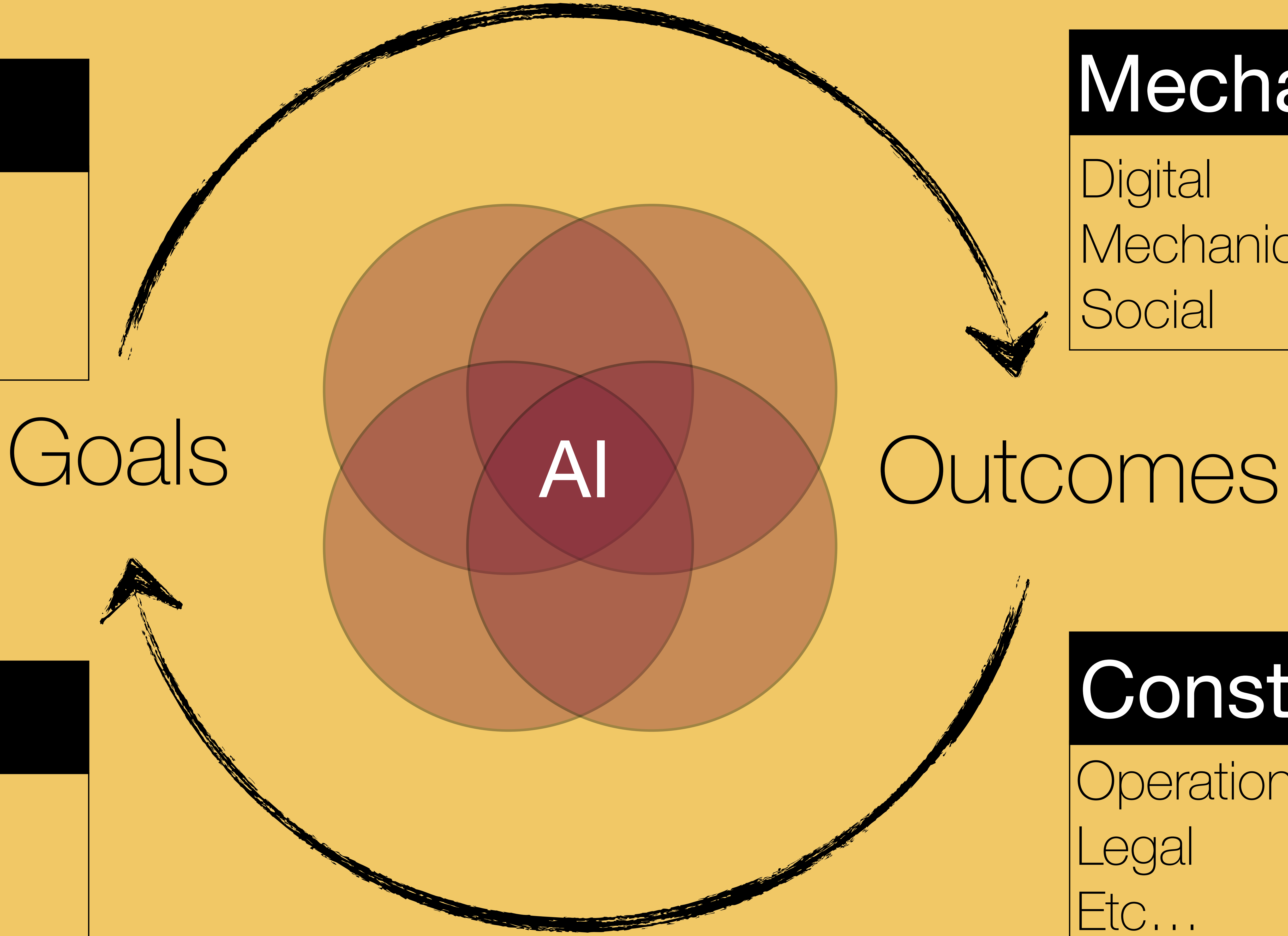
Harm ← **RISK** → Benefit

Data

Curated
Uncurated
Etc…
Closed
Open

“Knowledge”

Deterministic
Inferential
Etc…
Intuitive
Complex



Mechanisms

Digital
Mechanical
Social
Behavioral
Political
Etc…

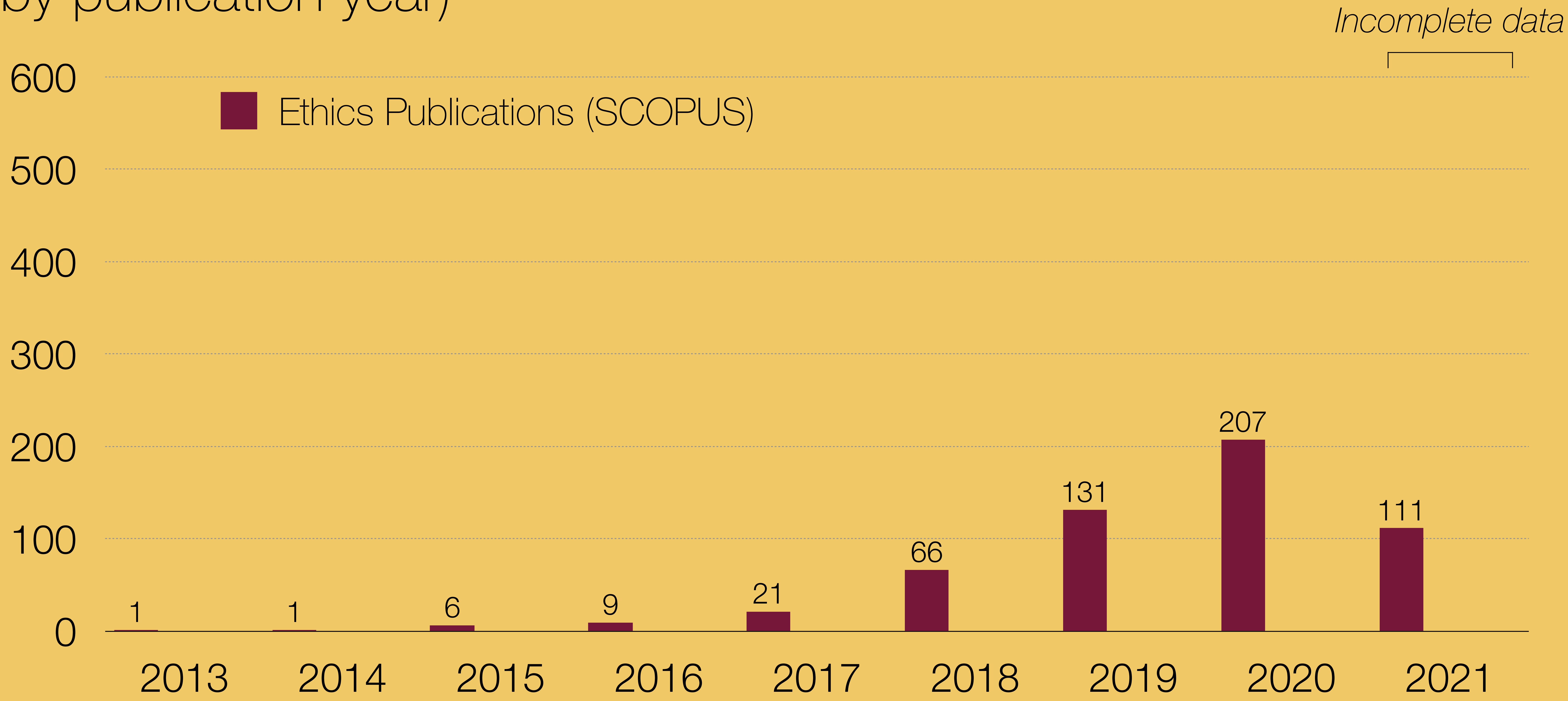
Constraints

Operational
Legal
Etc…
Institutional
Ethical

Wrong ← **ETHICS** → Right

AI Ethics vs Risk Publications

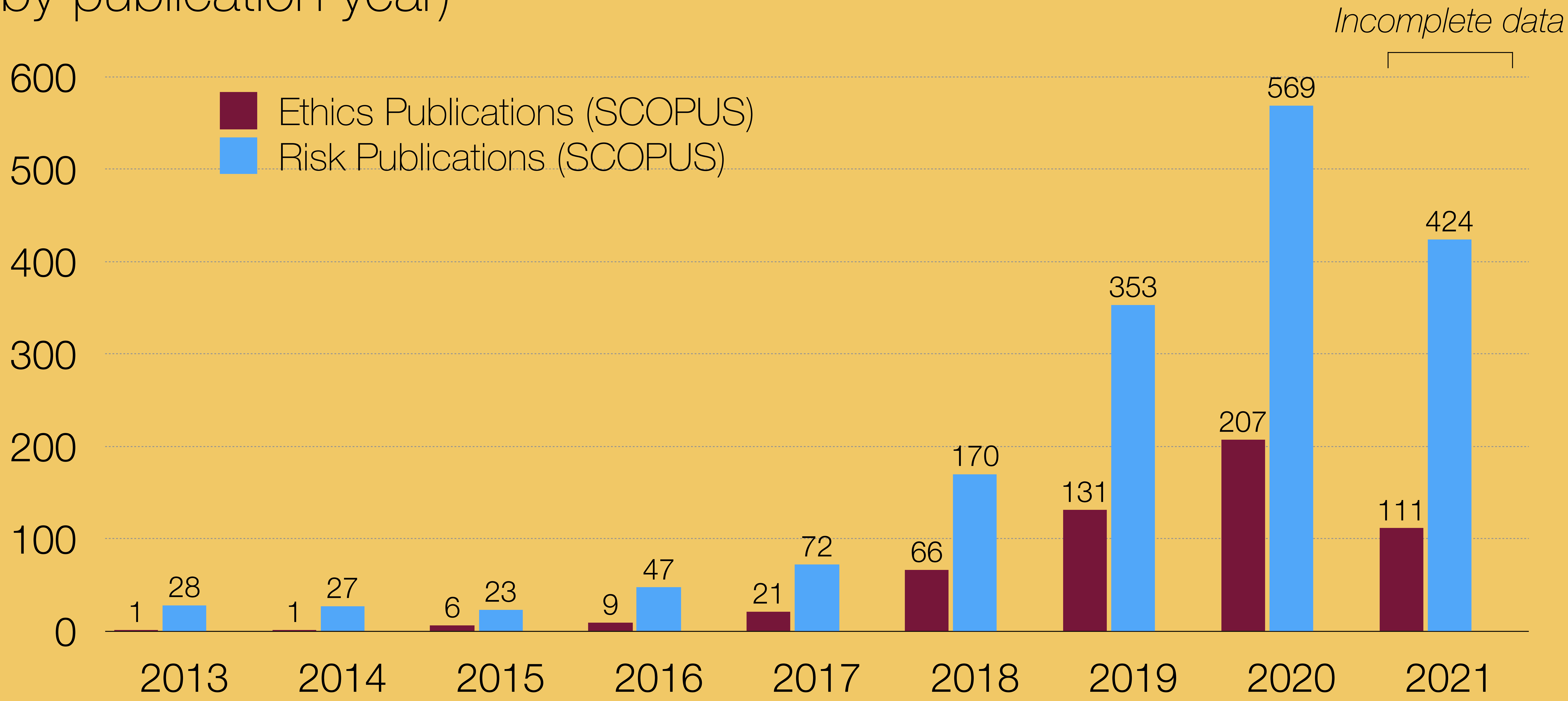
(by publication year)



SCOPUS search: TITLE (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning") AND (ethic*))
TITLE (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning") AND (risk*))

AI Ethics vs Risk Publications

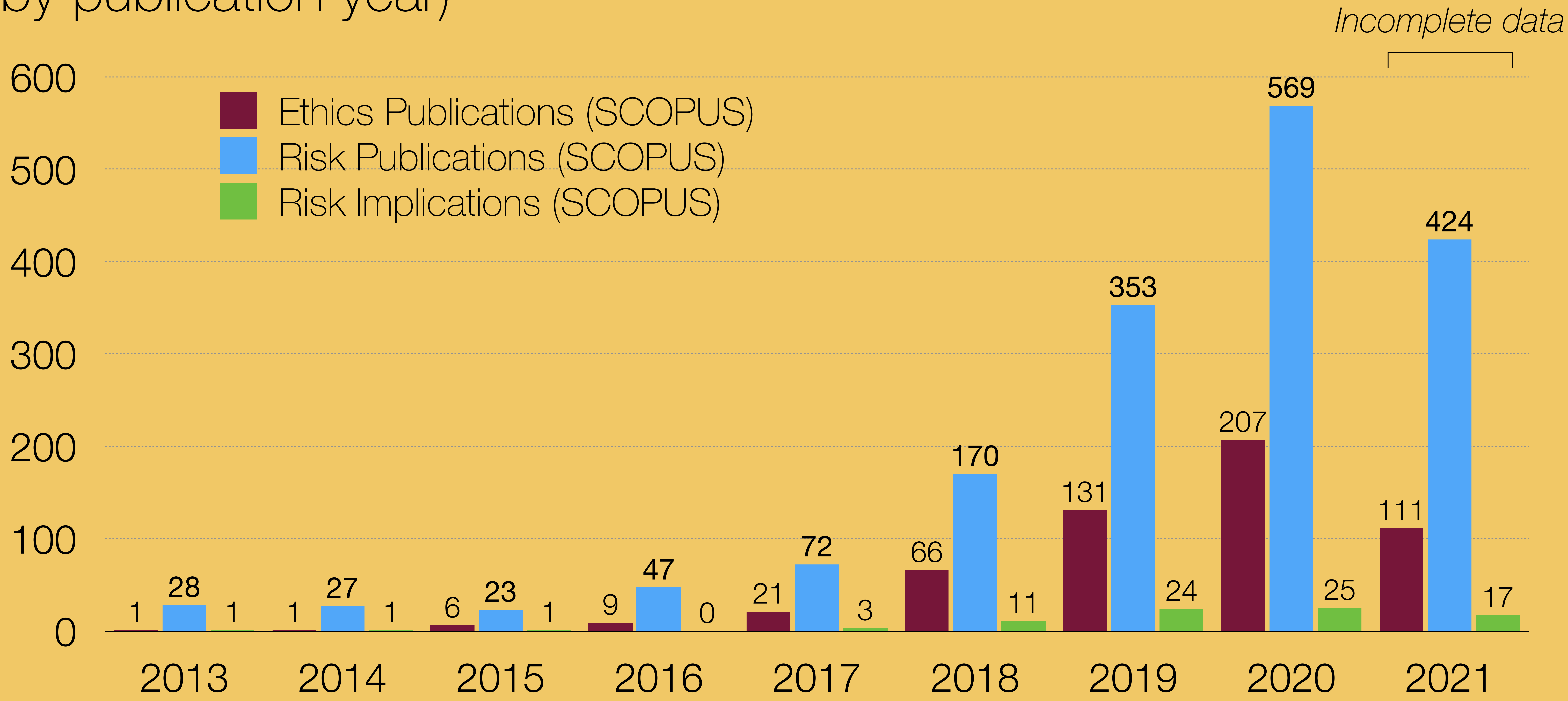
(by publication year)



SCOPUS search: TITLE (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning") AND (ethic*))
TITLE (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning") AND (risk*))

AI Ethics vs Risk Publications

(by publication year)



SCOPUS search: TITLE (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning") AND (ethic*))
TITLE (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning") AND (risk*))

RISK INNOVATION

RISK IS …

… the probability of harm occurring from an action or situation

INNOVATION IS …

… The process of translating an idea or invention into a good or service that creates value for which customers will pay

INNOVATION IS …

CREATIVITY

… The process of translating an idea or invention into a good or service that creates value for which customers will pay

PRODUCTS

VALUE

RISK INNOVATION IS …

A way of approaching risk that leads to new knowledge, understanding, and capabilities, and translates these into products, tools, or practices that protect and grow societal, environmental, economic, and other value

Maynard, A. D. Why We Need Risk Innovation Nature nanotechnology (2015) DOI:1038/nnano.2015.196

RISK INNOVATION IS …

CREATIVITY

A way of approaching risk that leads to new knowledge, understanding, and capabilities, and translates these into products, tools, or practices that protect and grow societal, environmental, economic, and other value

PRODUCTS

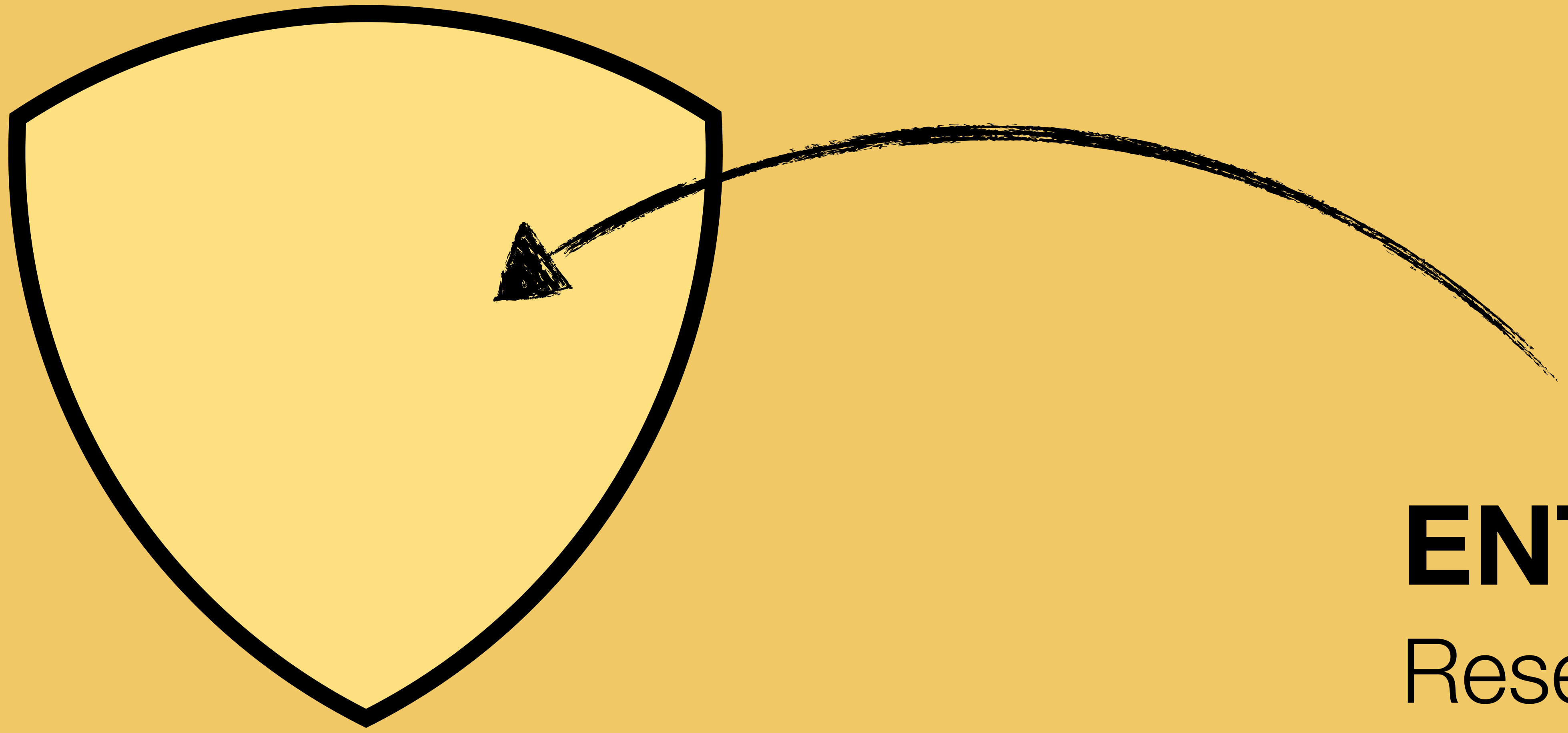
VALUE

Maynard, A. D. Why We Need Risk Innovation Nature nanotechnology (2015) DOI:1038/nnano.2015.196

RISK INNOVATION NEXUS

Connecting ethical and responsible innovation with value growth

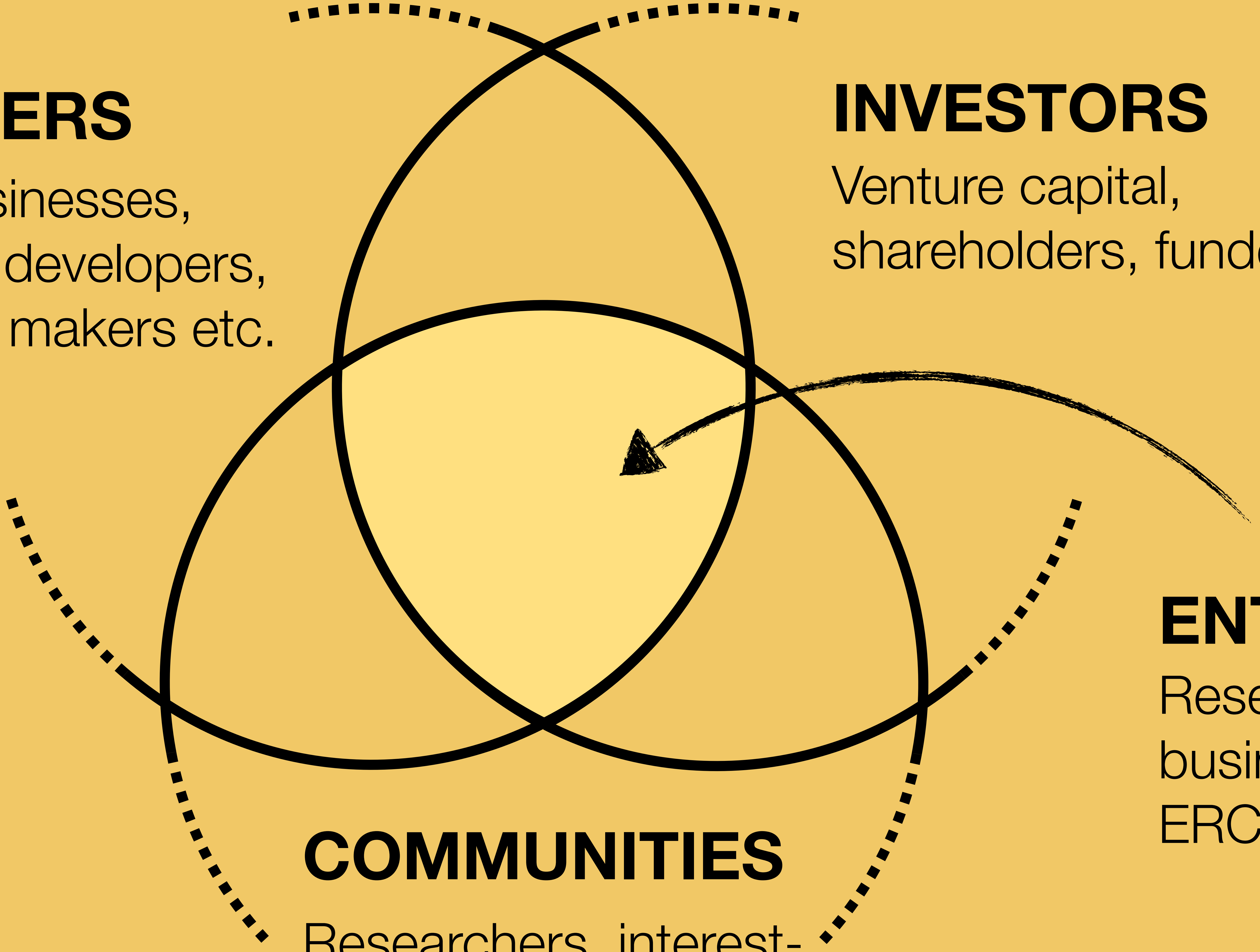
RiskInnovation.org



ENTERPRISE

Research program,
business, organization,
ERC, etc

STAKEHOLDERS



CUSTOMERS

Funders, businesses, researchers, developers, users, policy makers etc.

INVESTORS

Venture capital, shareholders, funders, etc.

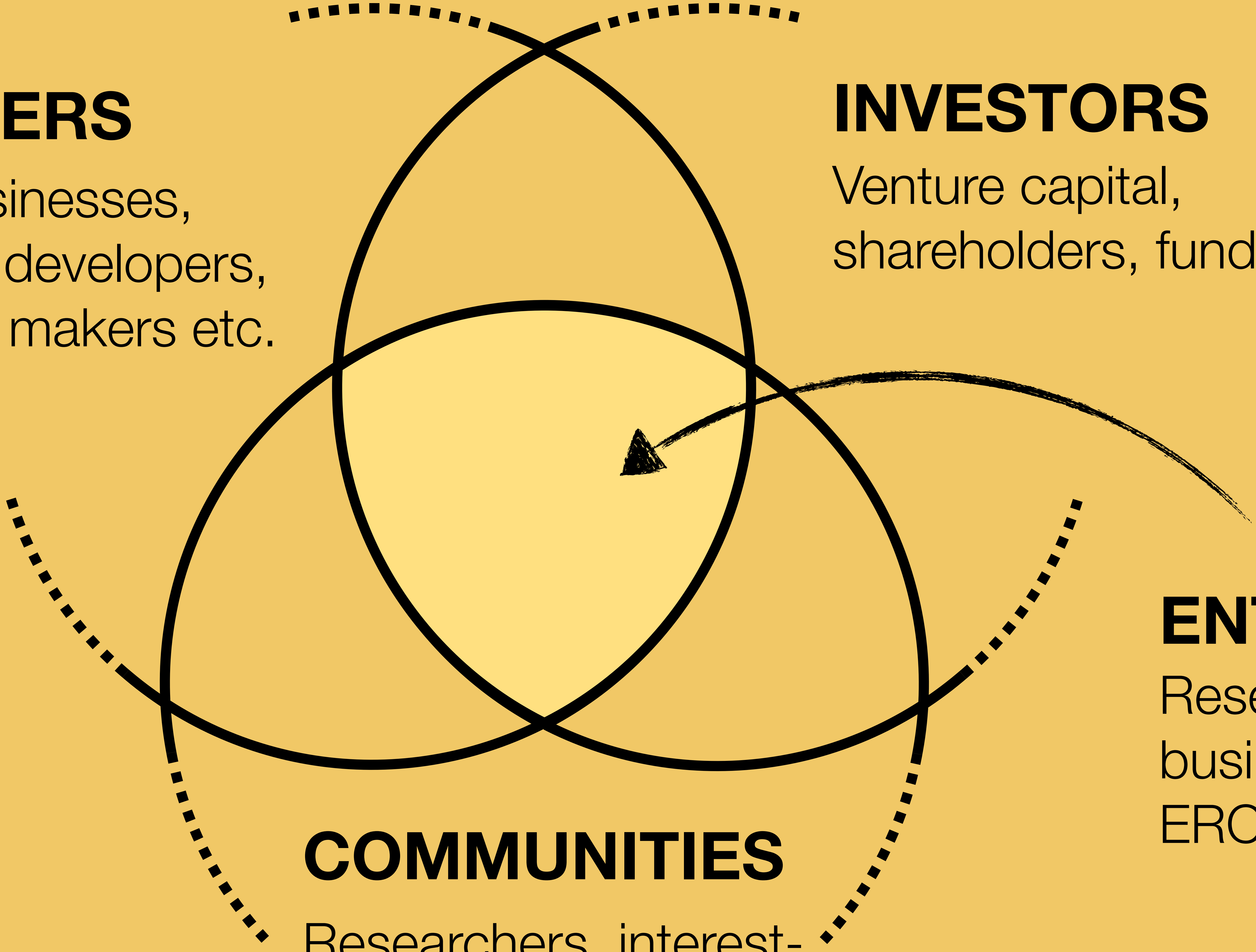
ENTERPRISE

Research program, business, organization, ERC, etc

COMMUNITIES

Researchers, interest-groups, civil society, publics, etc.

STAKEHOLDERS



CUSTOMERS

Funders, businesses, researchers, developers, users, policy makers etc.

INVESTORS

Venture capital, shareholders, funders, etc.

ENTERPRISE

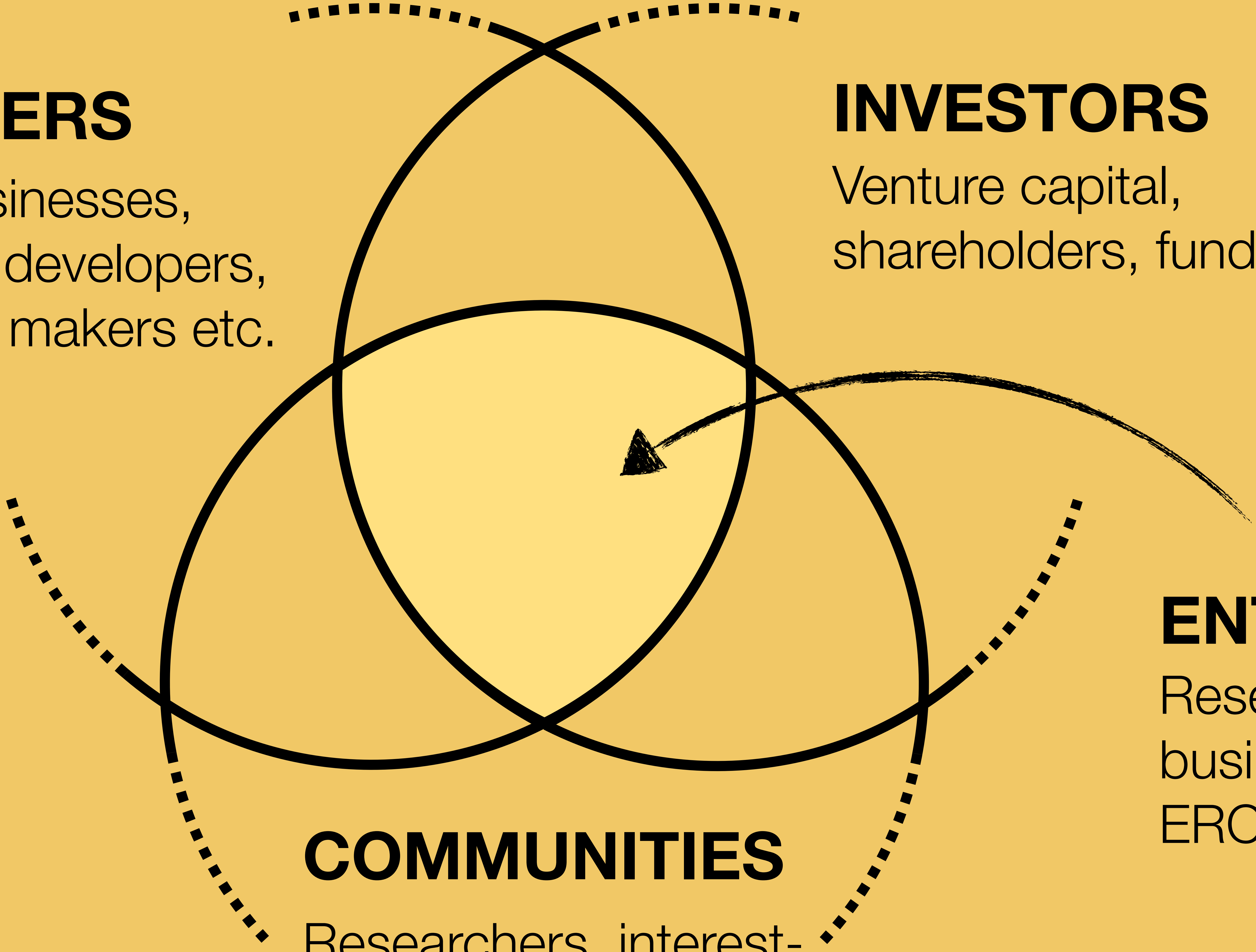
Research program, business, organization, ERC, etc

COMMUNITIES

Researchers, interest-groups, civil society, publics, etc.

VALUE

STAKEHOLDERS



CUSTOMERS

Funders, businesses, researchers, developers, users, policy makers etc.

INVESTORS

Venture capital, shareholders, funders, etc.

ENTERPRISE

Research program, business, organization, ERC, etc

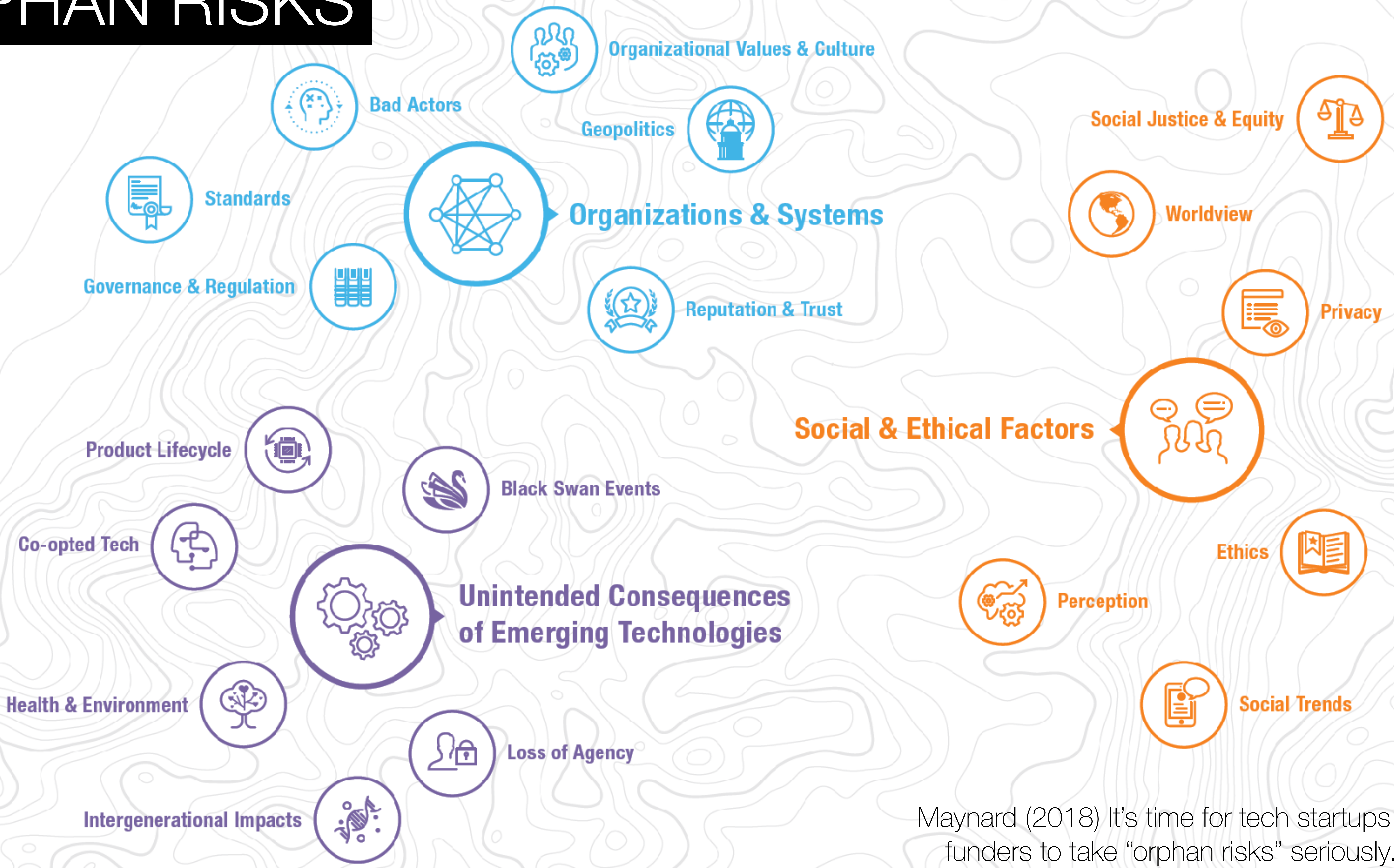
COMMUNITIES

Researchers, interest-groups, civil society, publics, etc.

THREATS

VALUE

ORPHAN RISKS



Organizations & Systems

- Organizational Values & Culture
- Bad Actors
- Geopolitics
- Standards
- Governance & Regulation
- Reputation & Trust

Social & Ethical Factors

- Social Justice & Equity
- Worldview
- Privacy
- Ethics
- Perception
- Social Trends

Unintended Consequences of Emerging Technologies

- Product Lifecycle
- Black Swan Events
- Co-opted Tech
- Health & Environment
- Loss of Agency
- Intergenerational Impacts

Maynard (2018) It's time for tech startups and their funders to take "orphan risks" seriously. Medium

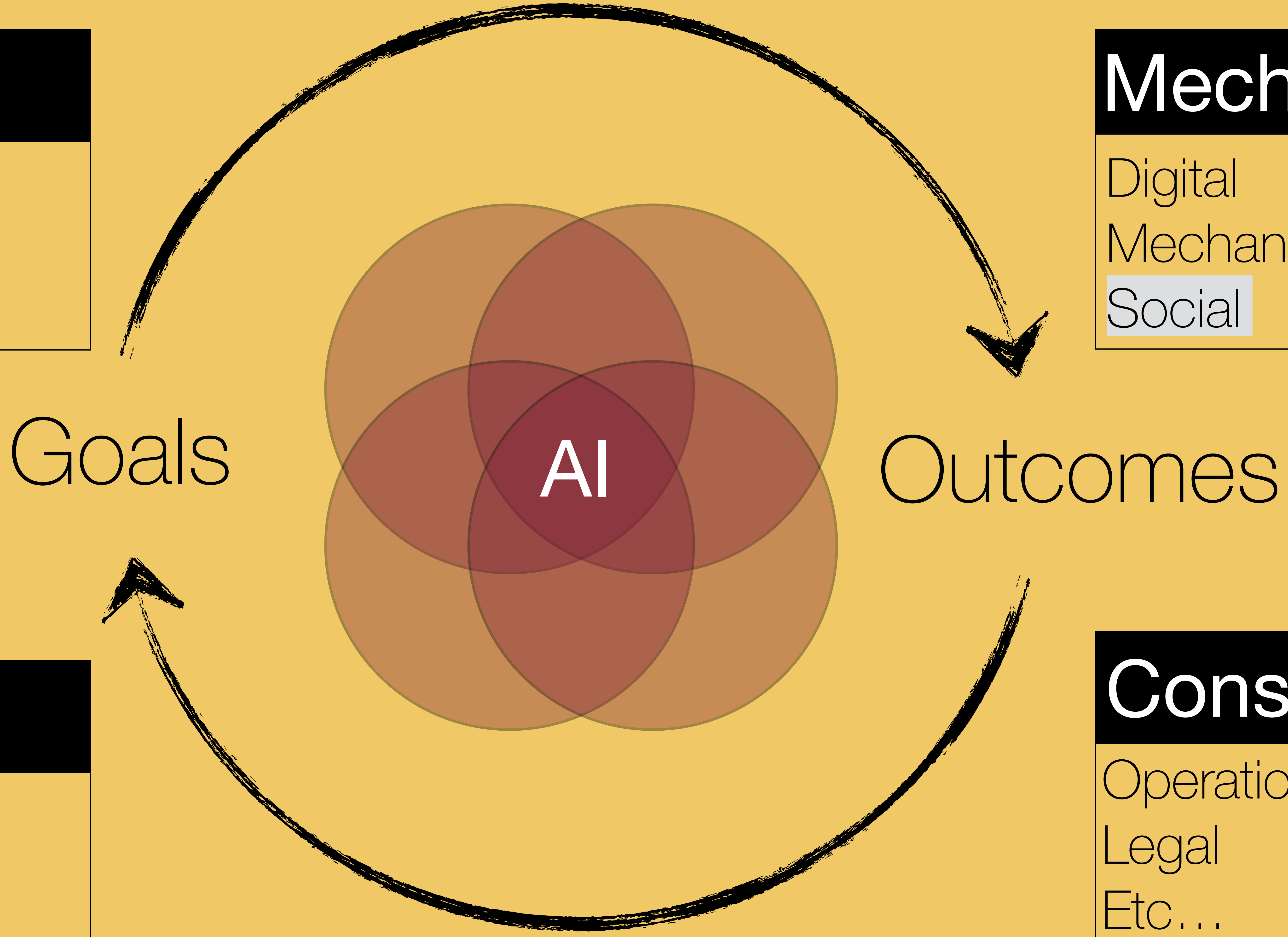
Hypothetical Example: Using AI-based social media agents to reach herd immunity

Data

- Curated
- Uncurated
- Etc…
- Closed
- Open

Mechanisms

- Digital
- Mechanical
- Social
- Behavioral
- Political
- Etc…



Knowledge

- Deterministic
- Inferential
- Etc…
- Intuitive
- Complex

Constraints

- Operational
- Legal
- Etc…
- Institutional
- Ethical

Orphan Risk Landscape

GOAL: Using AI-based social media agents to reach herd immunity (Private company)

VALUE

ENTERPRISE

INVESTORS

CUSTOMERS

COMMUNITIES

ORPHAN RISKS

Social & Ethical Factors

Unintended Consequences of Emerging Technologies

Organizations & Systems

Orphan Risk Landscape

GOAL: Using AI-based social media agents to reach herd immunity (Private company)

VALUE

ENTERPRISE

Behavior change at scale

Versatile tech platform

Profit

INVESTORS

Product that delivers on its promise

Trustworthiness

High return on investment

CUSTOMERS

Significantly reduced social & economic impacts of infectious agent

No legal/regulatory liability

Public support and recognition

COMMUNITIES

Autonomy

Transparency

Inclusion

ORPHAN RISKS

Social & Ethical Factors

Unintended Consequences of Emerging Technologies

Organizations & Systems

Orphan Risk Landscape

GOAL: Using AI-based social media agents to reach herd immunity (Private company)

VALUE

ENTERPRISE

Behavior change at scale

Versatile tech platform

Profit

INVESTORS

Product that delivers on its promise

Trustworthiness

High return on investment

CUSTOMERS

Significantly reduced social & economic impacts of infectious agent

No legal/regulatory liability

Public support and recognition

COMMUNITIES

Autonomy

Transparency

Inclusion

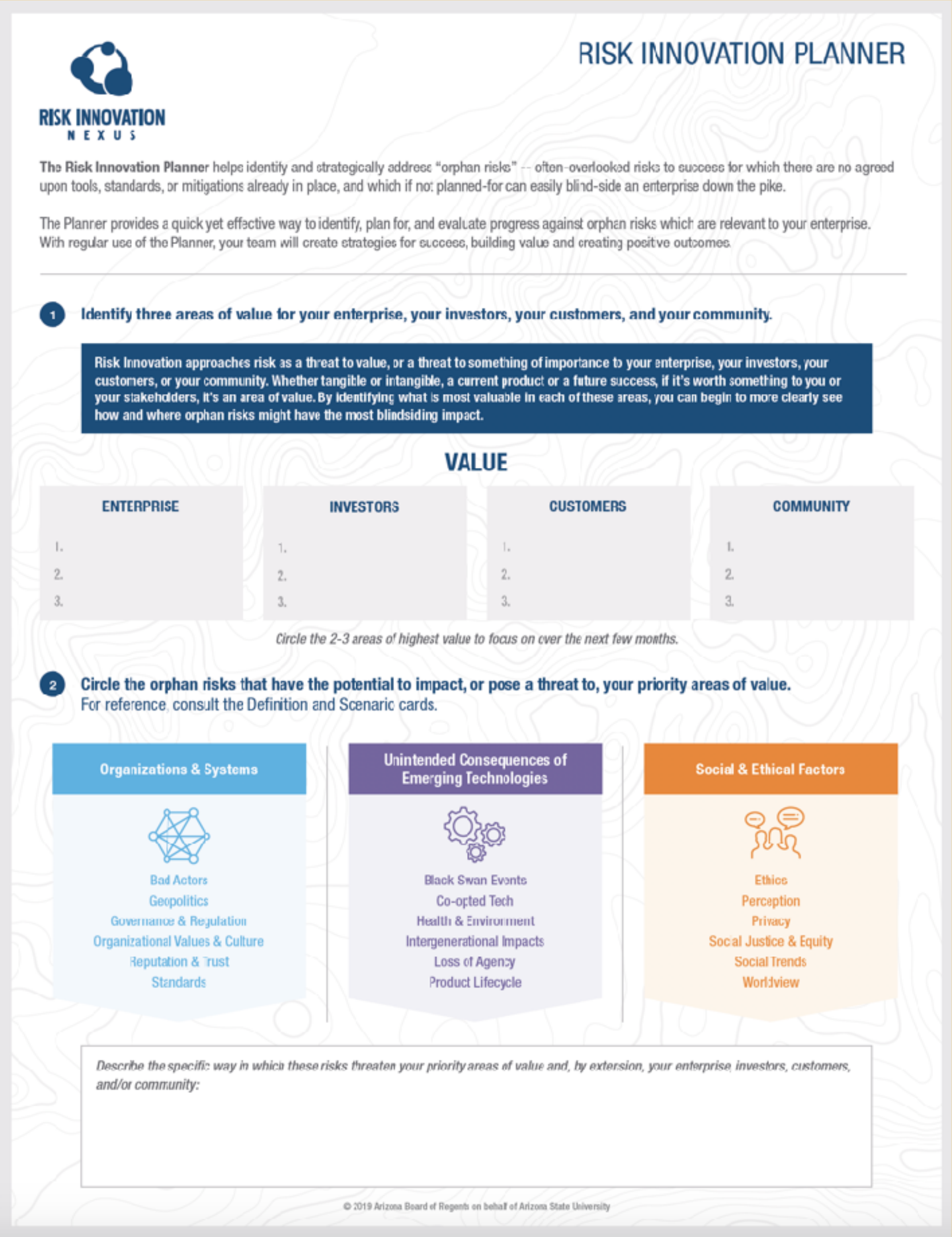
	Social & Ethical Factors	Unintended Consequences of Emerging Technologies	Organizations & Systems
ENTERPRISE	Privacy Social Trends Worldview	Loss of Agency	Governance & Regulation Organizational Values & Culture Standards
INVESTORS	Perception Ethics	Co-opted Tech Loss of Agency	Governance & Regulation Reputation & Trust Geopolitics
CUSTOMERS	Social Trends Perception Worldview	Health & Environment Co-opted Tech	Governance & Regulation Reputation & Trust Standards
COMMUNITIES	Privacy Social Justice & Equity Ethics	Co-opted Tech Loss of Agency	Reputation & Trust Bad Actors

RISK INNOVATION PLANNER

Short, iterative orphan risk audits and responses

One of a number of tools available through the Risk Innovation Nexus

Adaptable to multiple contexts



RISK INNOVATION PLANNER

RISK INNOVATION NEXUS

The Risk Innovation Planner helps identify and strategically address "orphan risks" -- often-overlooked risks to success for which there are no agreed upon tools, standards, or mitigations already in place, and which if not planned-for can easily blind-side an enterprise down the pike.

The Planner provides a quick yet effective way to identify, plan for, and evaluate progress against orphan risks which are relevant to your enterprise. With regular use of the Planner, your team will create strategies for success, building value and creating positive outcomes.

1 **Identify three areas of value for your enterprise, your investors, your customers, and your community.**

Risk Innovation approaches risk as a threat to value, or a threat to something of importance to your enterprise, your investors, your customers, or your community. Whether tangible or intangible, a current product or a future success, if it's worth something to you or your stakeholders, it's an area of value. By identifying what is most valuable in each of these areas, you can begin to more clearly see how and where orphan risks might have the most blindsiding impact.

VALUE

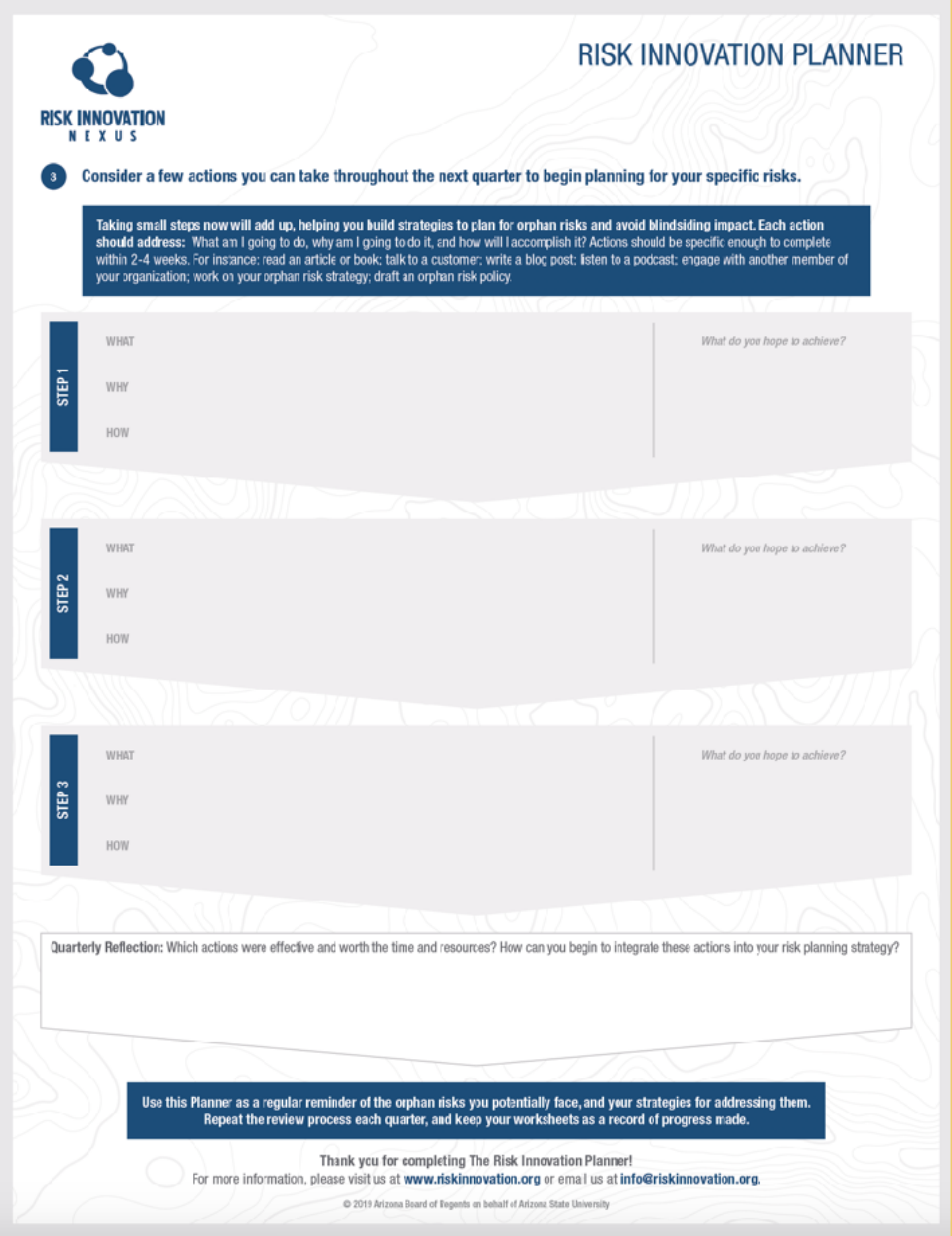
ENTERPRISE	INVESTORS	CUSTOMERS	COMMUNITY
1.	1.	1.	1.
2.	2.	2.	2.
3.	3.	3.	3.

Circle the 2-3 areas of highest value to focus on over the next few months.

2 **Circle the orphan risks that have the potential to impact, or pose a threat to, your priority areas of value.**
For reference, consult the Definition and Scenario cards.

Organizations & Systems	Unintended Consequences of Emerging Technologies	Social & Ethical Factors
Bad Actors	Black Swan Events	Ethics
Geopolitics	Co-opted Tech	Perception
Governance & Regulation	Health & Environment	Privacy
Organizational Values & Culture	Intergenerational Impacts	Social Justice & Equity
Reputation & Trust	Loss of Agency	Social Trends
Standards	Product Lifecycle	Worldview

Describe the specific way in which these risks threaten your priority areas of value and, by extension, your enterprise, investors, customers, and/or community:



RISK INNOVATION PLANNER

RISK INNOVATION NEXUS

3 **Consider a few actions you can take throughout the next quarter to begin planning for your specific risks.**

Taking small steps now will add up, helping you build strategies to plan for orphan risks and avoid blindsiding impact. Each action should address: What am I going to do, why am I going to do it, and how will I accomplish it? Actions should be specific enough to complete within 2-4 weeks. For instance: read an article or book; talk to a customer; write a blog post; listen to a podcast; engage with another member of your organization; work on your orphan risk strategy; draft an orphan risk policy.

STEP 1

WHAT

WHY

HOW

What do you hope to achieve?

STEP 2

WHAT

WHY

HOW

What do you hope to achieve?

STEP 3

WHAT

WHY

HOW

What do you hope to achieve?

Quarterly Reflection: Which actions were effective and worth the time and resources? How can you begin to integrate these actions into your risk planning strategy?

Use this Planner as a regular reminder of the orphan risks you potentially face, and your strategies for addressing them. Repeat the review process each quarter, and keep your worksheets as a record of progress made.

Thank you for completing The Risk Innovation Planner!
For more information, please visit us at **www.riskinnovation.org** or email us at **info@riskinnovation.org.**

https://riskinnovation.org/services/risk-innovation-planner/

Key threats to value that need to be addressed if social good is to be realized

- Ethics
- Social Trends
- Perception
- Privacy
- Worldview
- Governance & Regulation
- Loss of Agency
- Reputation & Trust
- Co-opted Tech

Loss of Agency

People are highly manipulatable

Human-human manipulation is constrained by a level(ish) playing field

Machine-human manipulation is not constrained in the same ways

Machines can potentially be taught (or learn) to take advantage of human heuristics and cognitive biases

This might be a beneficial thing for the future of humanity …

… or it could be really harmful!

Loss of Agency

How do we navigate a machine-mediated future of cognitive and behavioral asymmetry?

People are highly manipulatable

Human-human manipulation is constrained by a level(ish) playing field

Machine-human manipulation is not constrained in the same ways

Machines can potentially be taught (or learn) to take advantage of human heuristics and cognitive biases

This might be a beneficial thing for the future of humanity …

… or it could be really harmful!

KEY TAKEAWAY:

Socially beneficial and responsible development and use of “AI” requires new thinking around value and risk, as well as ethics

ASU College of Global Futures
Arizona State University

PROFESSOR ANDREW MAYNARD

Associate Dean
College of Global Futures
Arizona State University

Email: andrew.maynard@asu.edu
Twitter: @2020science
Medium: medium.com/edge-of-innovation
Web: andrewmaynard.net