

# Ethics of Artificial (Narrow) Intelligence

Nicholas Kalogirou, P.Eng | July 16 2020





"The real problem of humanity is the following: we have paleolithic emotions; medieval institutions; and god-like technology."

- E.O. Wilson, Biologist





Why does AI Ethics matter?

What are we talking about?

A dashed white crosshair graphic consisting of a vertical line and a horizontal line intersecting at the center of the page.

What are we talking about?

A nested systems model

What are we talking about?

A nested systems model

The current landscape

What are we talking about?

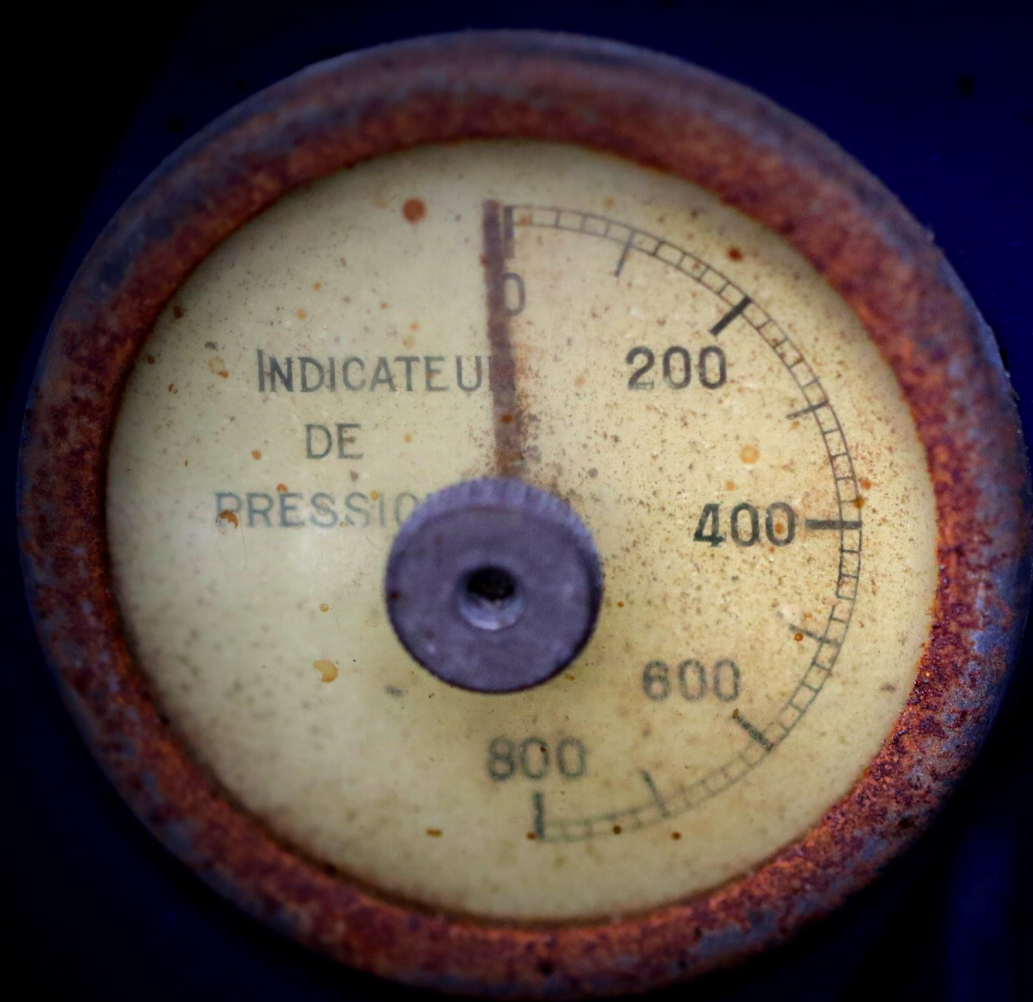
A nested systems model

The current landscape

What can we do?



PART 1  
WHAT ARE WE TALKING ABOUT



"In their area of practice, engineers shall hold paramount the health, safety, and welfare of the public, and have regard for the environment."

Guideline for Ethical Practice v2.2

Association of Professional Engineers and Geoscientists of Alberta (APEGA)

How do you know whether you are doing  
more good than harm?

**ethics** is the *ongoing* study, development, and application of moral reasoning

today's talk: the ethics of artificial **narrow** intelligence

PART 2  
NESTED SYSTEMS



model





A Venn diagram consisting of two overlapping circles. The left circle is white with a white outline and contains the word 'model'. The right circle is teal with a teal outline and contains the word 'modeller'. The two circles overlap in the center.

model

modeller

A Venn diagram on a dark gray background. It consists of three nested shapes: a small white circle on the left containing the word 'model', a larger teal oval in the middle containing the word 'modeller', and a large orange oval on the right containing the word 'organization'. The 'model' circle is entirely within the 'modeller' oval, and the 'modeller' oval is entirely within the 'organization' oval.

model

modeller

organization

A Venn diagram on a dark gray background. It consists of three nested shapes: a small white circle on the left, a larger teal oval in the middle, and a large orange oval on the right. The white circle is entirely contained within the teal oval. The teal oval is entirely contained within the orange oval. The word 'model' is centered in the white circle. The word 'modeller' is centered in the teal oval. The word 'organization' is centered in the orange oval. The phrase 'get paid' is positioned between the teal oval and the orange oval, and the phrase 'profit' is positioned between the orange oval and the right edge of the diagram.

model

modeller

organization

*get paid*

*profit*

A Venn diagram on a dark gray background. It consists of three nested shapes: a small white circle on the left, a larger teal oval in the middle, and a large orange oval on the right. The white circle is entirely contained within the teal oval, which is in turn contained within the orange oval. The word 'model' is centered in the white circle. The words 'modeller' and 'opp cost' are centered in the teal oval. The words 'organization', 'project', and 'failure' are centered in the orange oval.

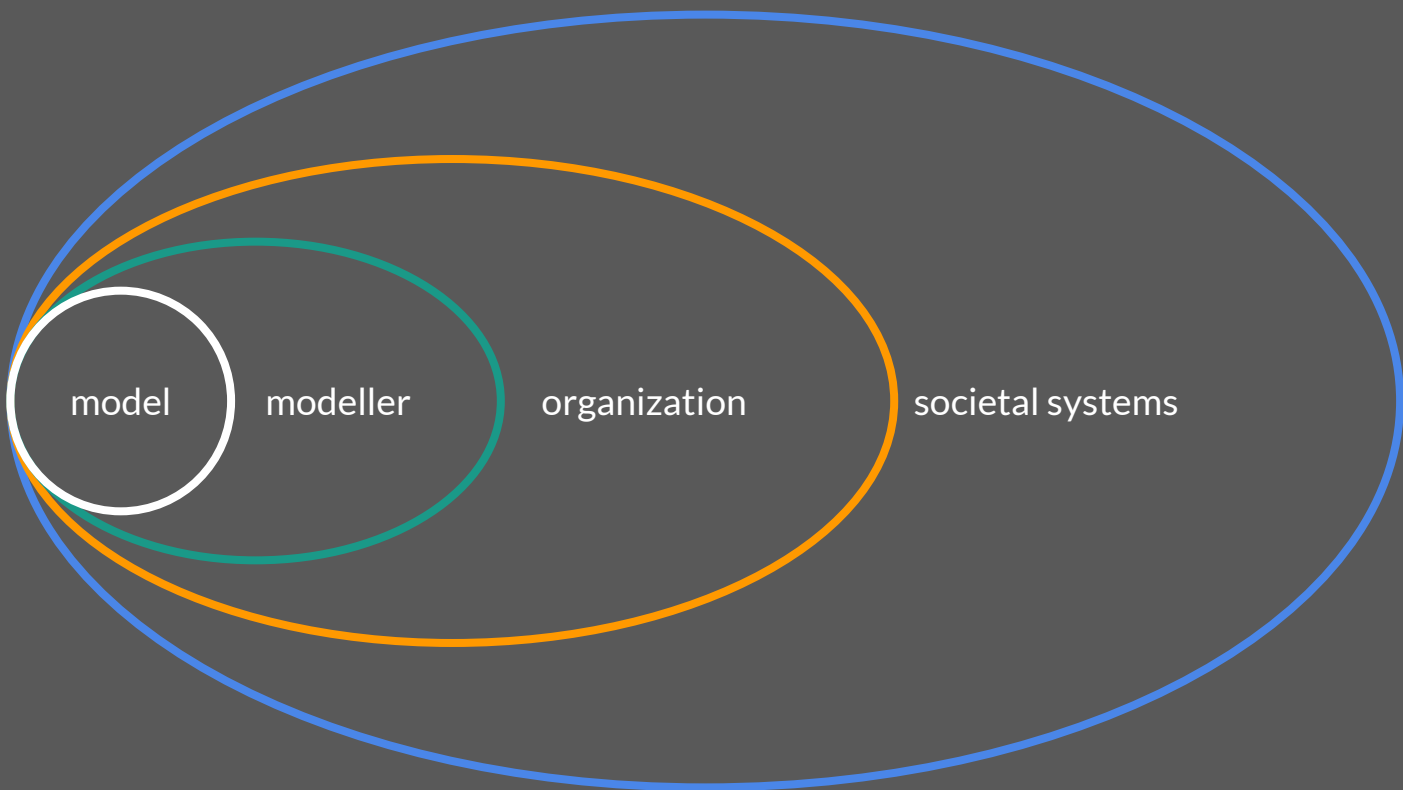
model

modeller

*opp cost*

organization

*project  
failure*

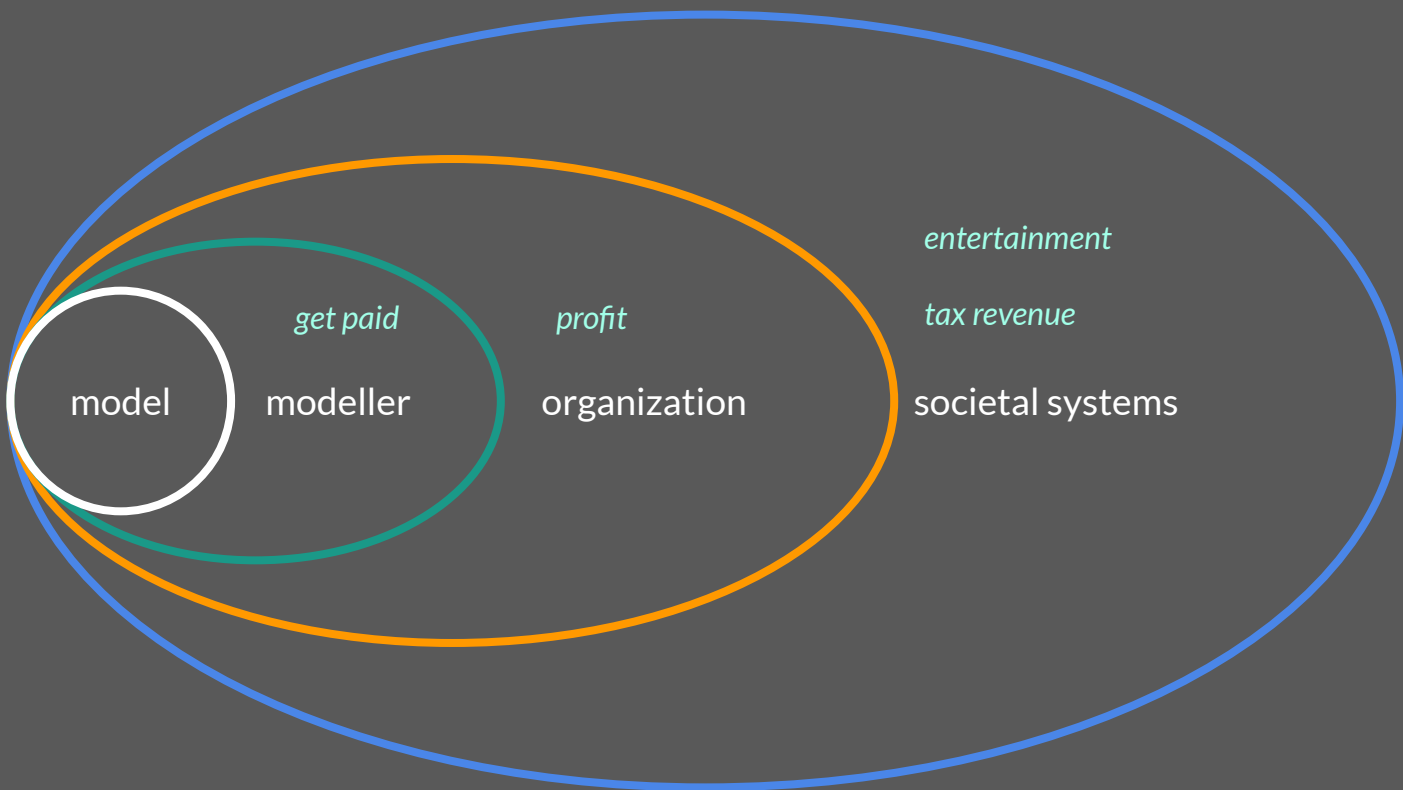


model

modeller

organization

societal systems



model

modeller

*get paid*

organization

*profit*

societal systems

*tax revenue*

*entertainment*

model

modeller

*opp cost*

organization

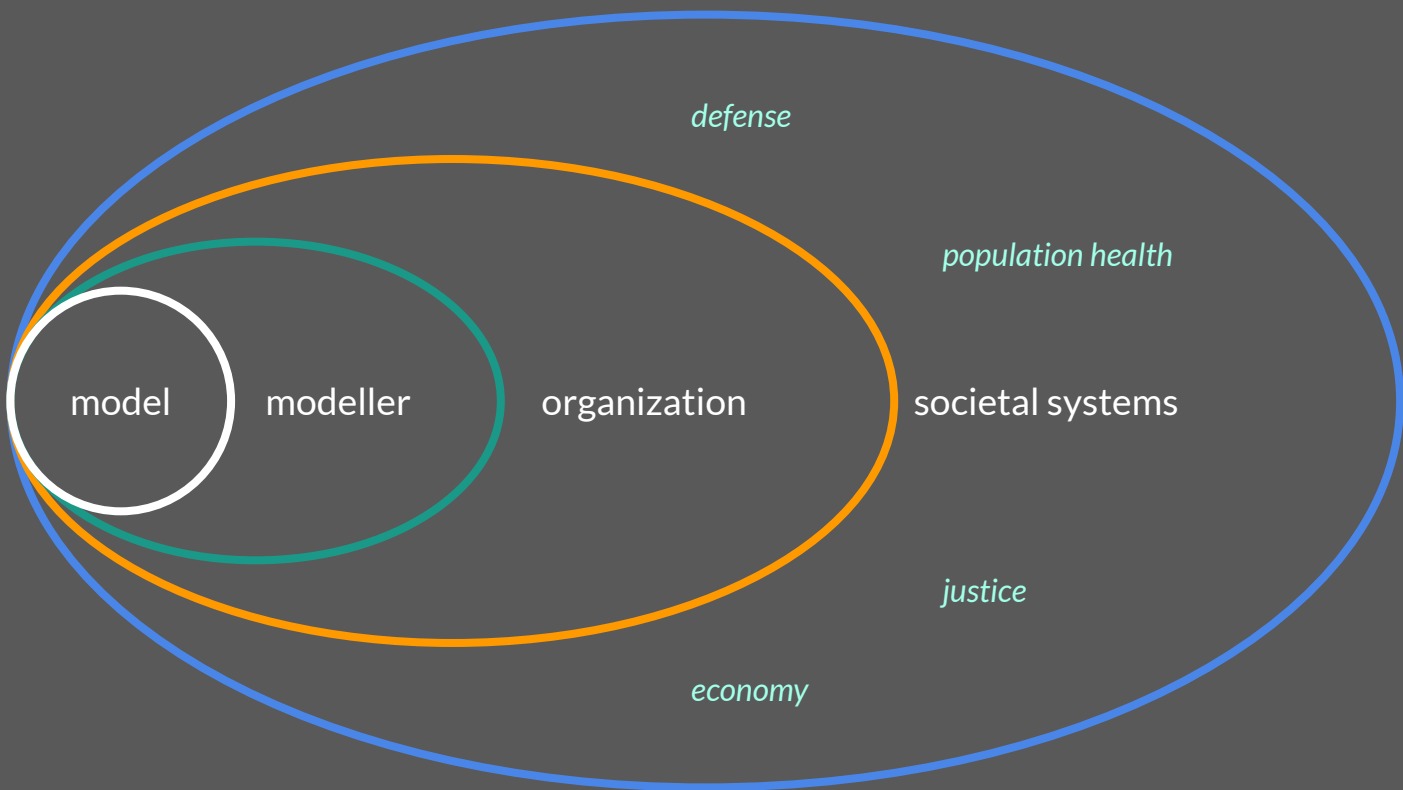
*project  
failure*

societal systems

*vulnerable population  
health + well being*

*difficult to escape*





model

modeller

organization

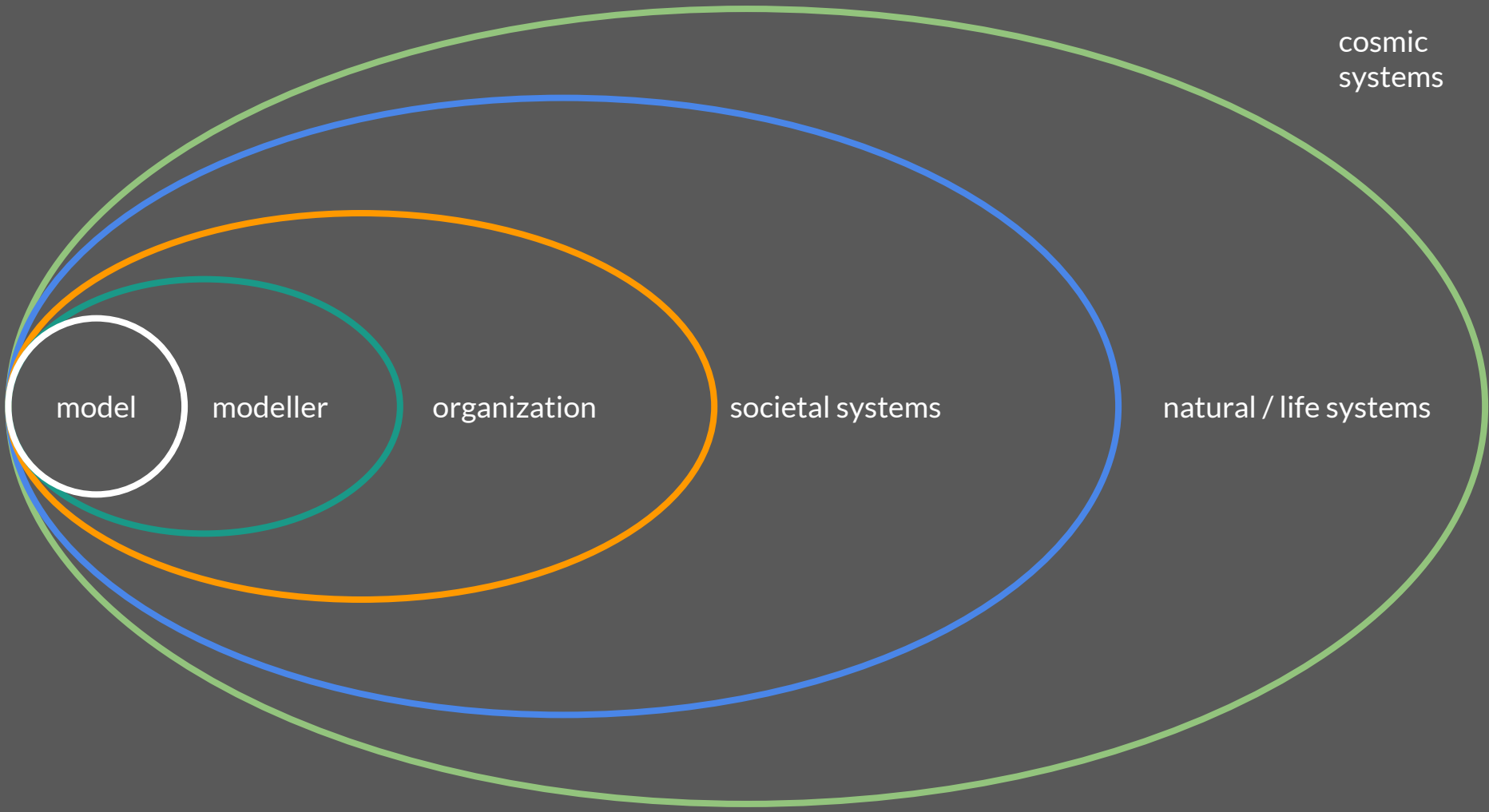
societal systems

*defense*

*population health*

*justice*

*economy*



model

modeller

organization

societal systems

natural / life systems

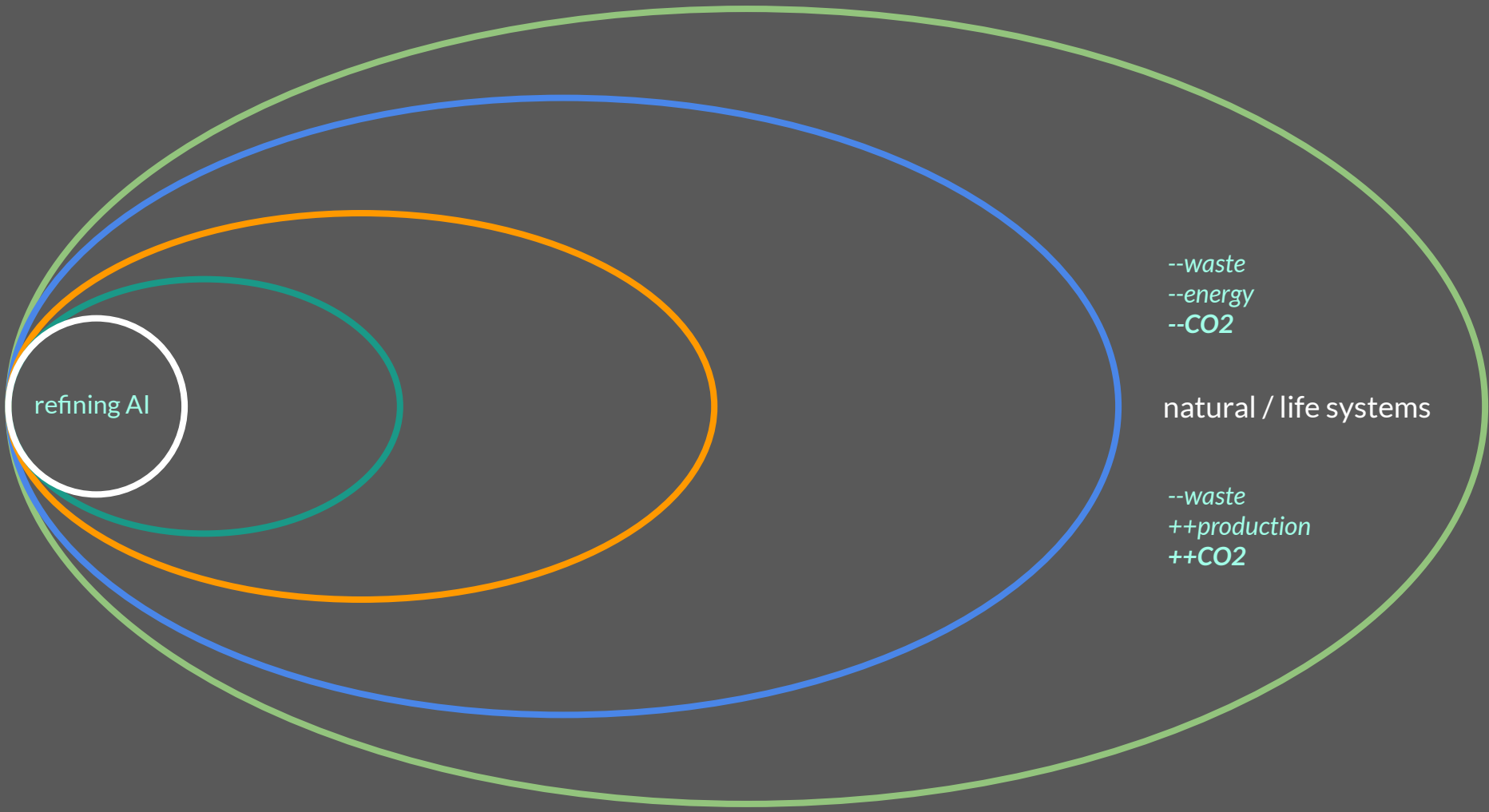
cosmic  
systems



refining AI

--waste  
--energy  
--CO2

natural / life systems

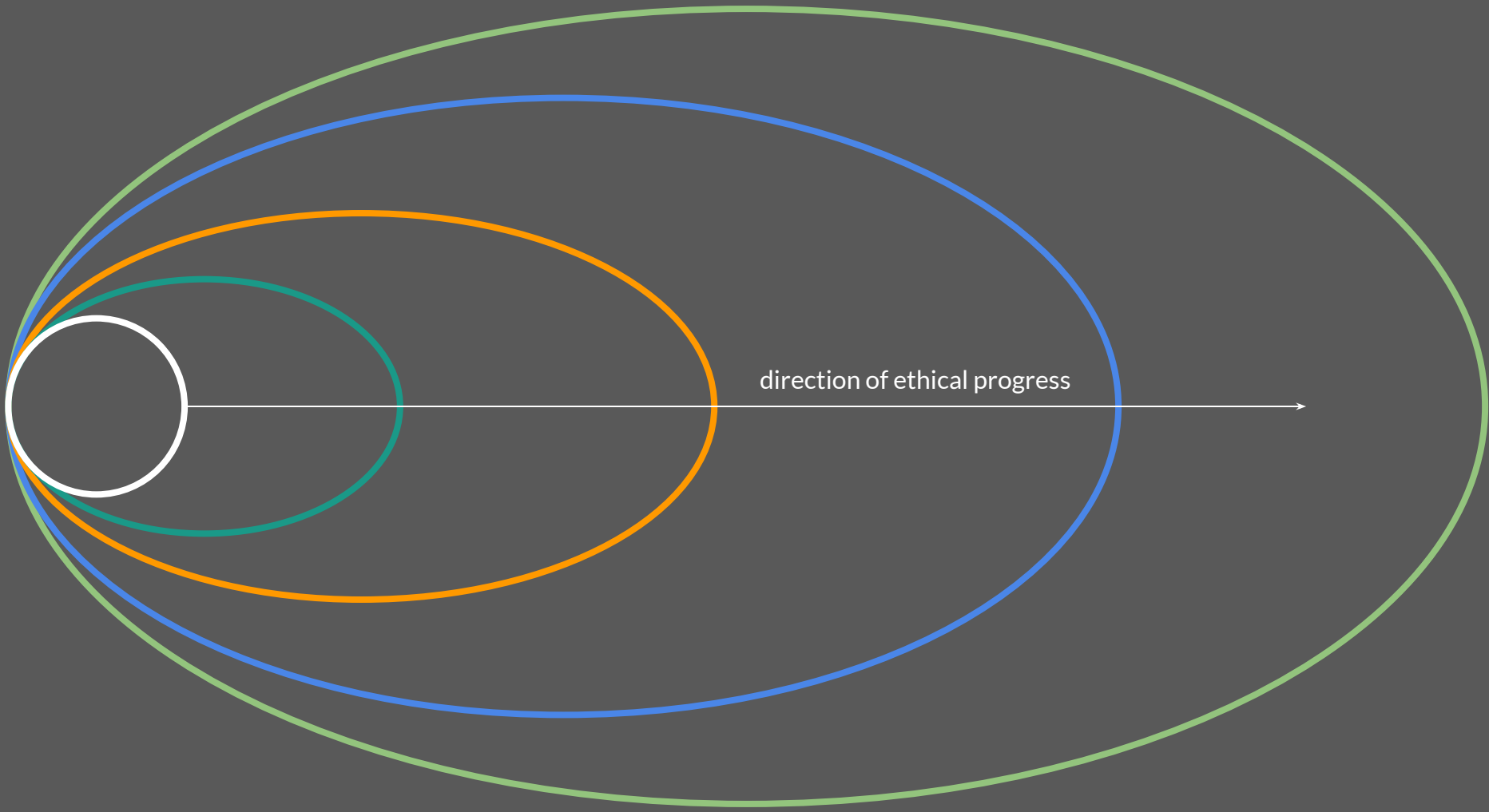


refining AI

--waste  
--energy  
--CO2

natural / life systems

--waste  
++production  
++CO2

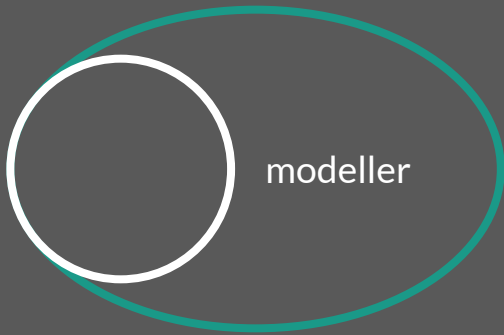


direction of ethical progress

# "Models are opinions embedded in mathematics"

- Cathy O'Neill , "Weapons of Math Destruction"





What is the boundary of my caring?

What are my values?

What responsibility do I accept?

What do I have the courage to question?

How do I verify my own knowledge?



The diagram consists of three nested shapes on the left: a small white circle, a teal oval, and a large light green oval. The word 'modeller' is written in white text to the right of the white circle. In the center, the text 'look out' and 'look in' is displayed. To the right of this text are two horizontal white arrows: the top one points right and the bottom one points left.

modeller

look out  
look in



PART 3  
THE CURRENT LANDSCAPE

**ACADEMICS**

**DATA  
PRACTITIONERS**

**TECHNOLOGISTS**

**GOVERNMENTS**

**ACADEMICS**

**NON-PROFITS**

**DATA  
PRACTITIONERS**

**ACADEMIC  
INSTITUTIONS**

**TECHNOLOGISTS**

**CORPORATIONS**



		frameworks checklists	Algorithmic Impact Assessment
	GOVERNMENTS	interdisciplinary research	information ethics
ACADEMICS	NON-PROFITS	technical tools	XAI
DATA PRACTITIONERS	ACADEMIC INSTITUTIONS	laws / regulation	GDPR
TECHNOLOGISTS	CORPORATIONS	audits / deployment	governance
		guidelines best practices	MILA Statement

		frameworks checklists	Algorithmic Impact Assessment
	GOVERNMENTS	interdisciplinary research	information ethics
ACADEMICS	NON-PROFITS	technical tools	XAI
DATA PRACTITIONERS	ACADEMIC INSTITUTIONS	laws / regulation	GDPR
TECHNOLOGISTS	CORPORATIONS	audits / deployment	governance
		guidelines best practices	MILA Montreal Declaration

# a review of AI ethics guidelines

## major themes

accountability

privacy

fairness

transparency

safety and  
security

common good

human  
oversight

explainability

inclusion  
social cohesion



80% of guidelines include...

**accountability**

**privacy**

**fairness**

**significant technical efforts**

# Where are the gaps?

## seeing ethical AI mainly as a technical problem

*other approaches like collective action - self organization, right incentives, right public policy*

# Where are the gaps?

## violating ethics standards and codes have no consequences

*technology outpaces the law | no professional bodies or regulations to reinforce behavior*

# Where are the gaps?

## vague guidelines and little focus on self development

*reading guidelines tend to have no influence | developing our own self-responsibility and caring*

# Where are the gaps?

**skipping ethics for profit and efficiency**

*lack of time and resources for broader questioning*

## Where are the gaps?

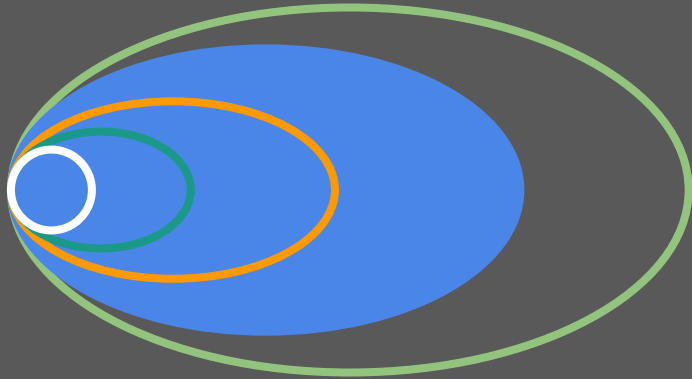
seeing ethical AI mainly as a technical problem

violating ethics standards and codes have no consequences

vague guidelines and little focus on self development

skipping ethics for profit and efficiency

# harmful consequences of AI - social trends



## INEQUALITY

Worker / Employer Inequality  
Widening Socioeconomic Gaps  
Political / Democratic Disruption

## LOSS OF LIBERTY

Privacy IS Liberty  
Increased control by corporations and  
government through surveillance

## DECAY OF TRUTH

Bad actors manipulate AI systems to  
control the narrative for narrow gain  
Intentional attacks on our shared  
understanding

PART 4  
WHAT CAN WE DO



# DEVELOP SELF

*Widen your caring  
not just technicals!*

WHAT CAN  
WE DO

# DEVELOP SELF

*Widen your caring  
not just technicals!*

WHAT CAN  
WE DO

# DISCUSS VALUES

*Openly discuss gap  
between values + action*

## DEVELOP SELF

*Widen your caring  
not just technicals!*

## QUESTION BROADLY

*Who / what are we empowering?  
not just cost efficiency  
Keep asking - write it down*

## WHAT CAN WE DO

## DISCUSS VALUES

*Openly discuss gap  
between values + action*

## DEVELOP SELF

*Widen your caring  
not just technicals!*

## QUESTION BROADLY

*Who / what are we empowering?  
not just cost efficiency  
Keep asking - write it down*

## WHAT CAN WE DO

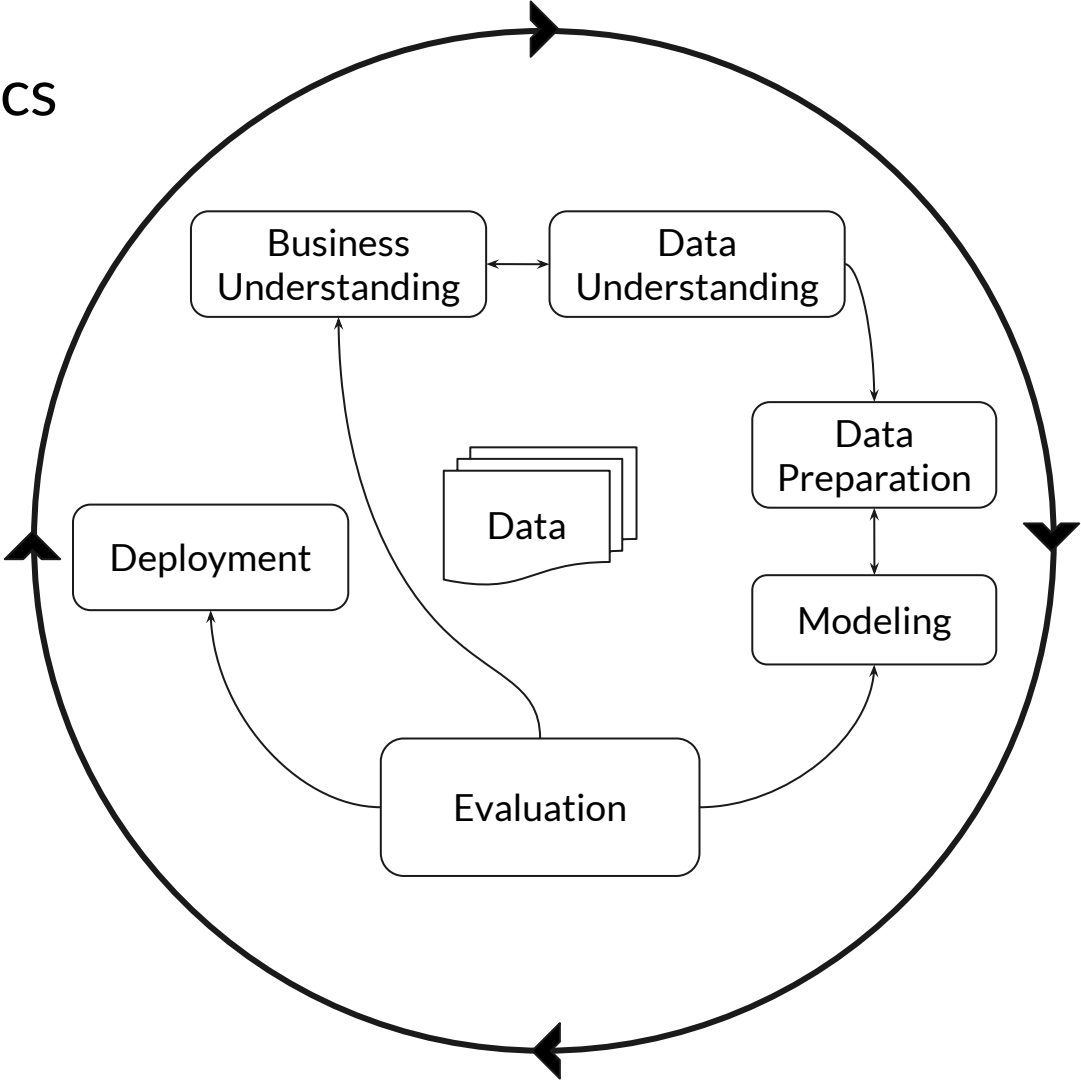
## DISCUSS VALUES

*Openly discuss gap  
between values + action*

## PUBLIC PRESSURE

*Be an active citizen  
Change the system, not just the  
individual / organization*

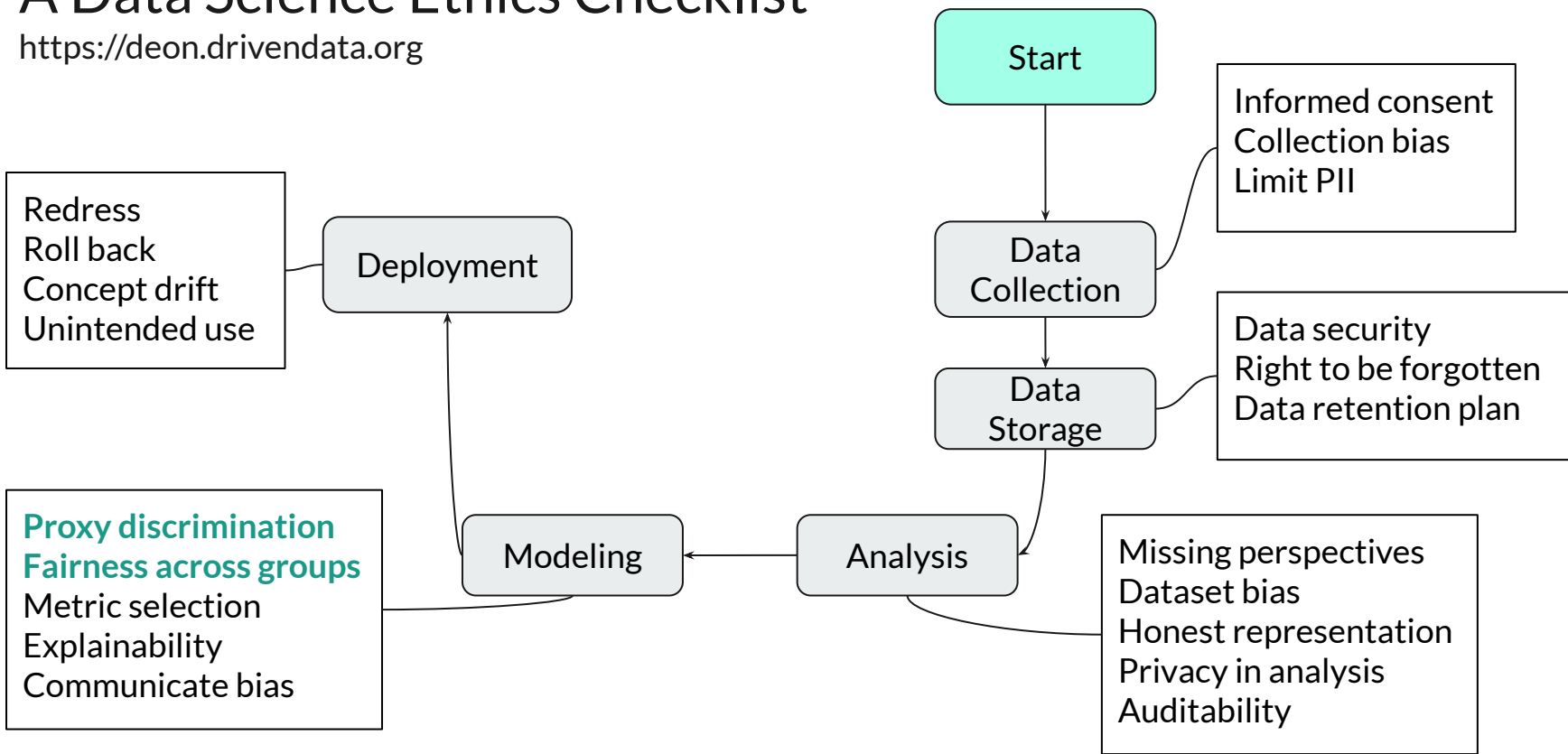
Data Analytics  
Lifecycle  
CRISP-DM



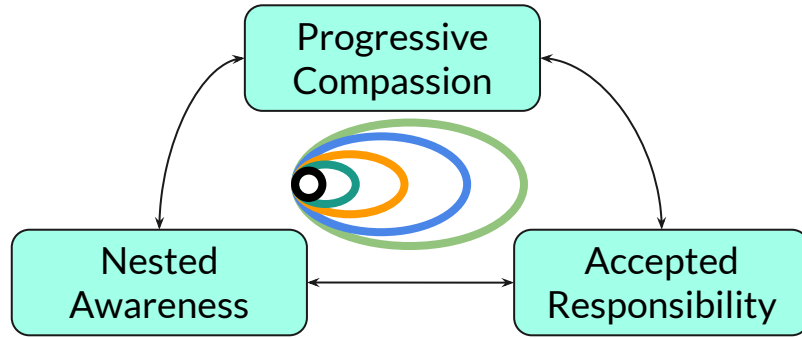


# A Data Science Ethics Checklist

<https://deon.drivendata.org>



# Nick's Ethics Process v0.1



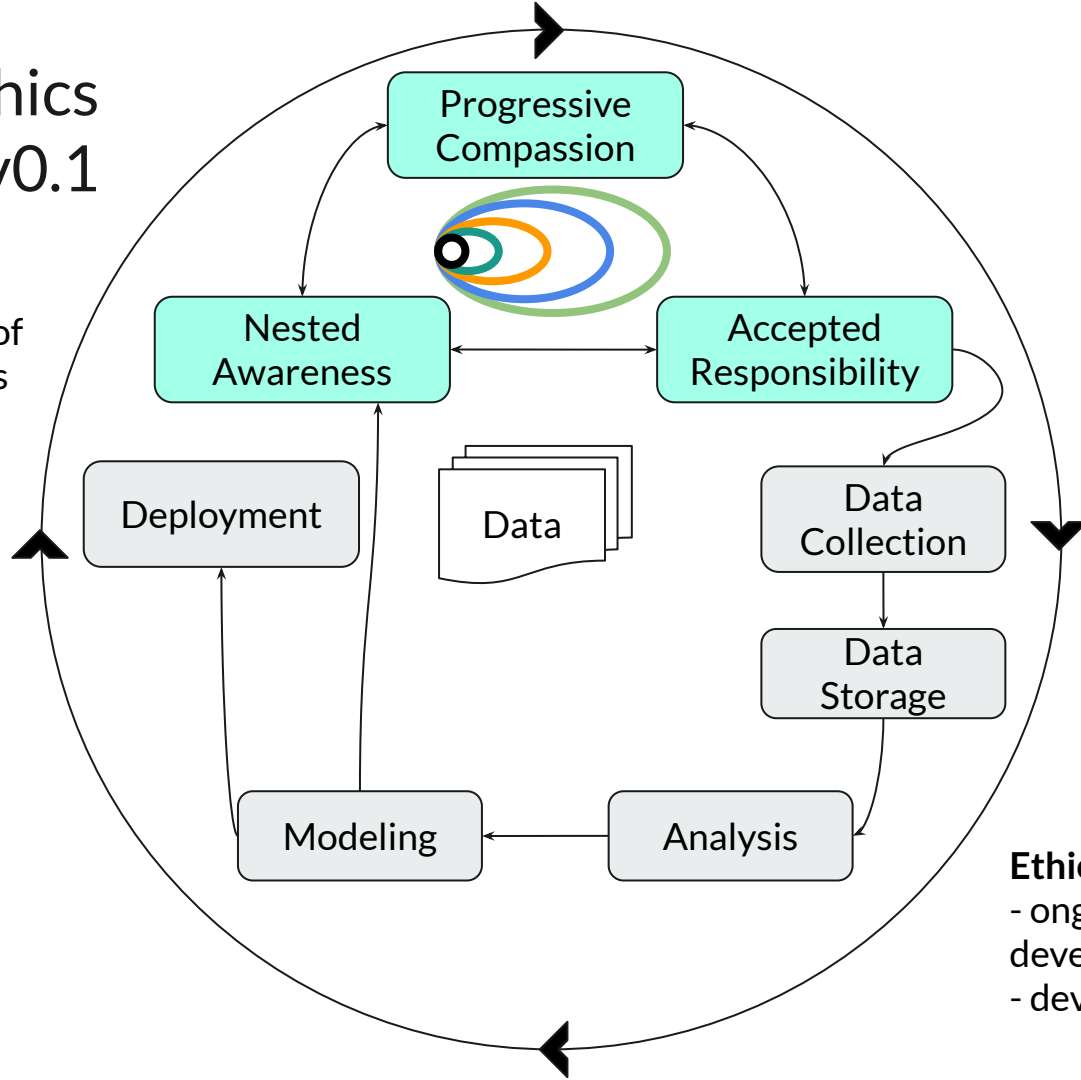
Explore impact of broader systems

Set progressive caring objective

Choose to act

# Nick's Ethics Process v0.1

Explore impact of broader systems



Set progressive caring objective

Choose to act

## Ethics Feedback Loop

- ongoing study, revision, development of moral reasoning
- develop oversight



**“We make our world significant  
by the courage of our questions,  
and the depth of our answers”**

**- Carl Sagan**



# References

Alberta Boiler Safety Association. (n.d.). About Us. History of ABSA & Heritage: Heritage. <https://www.absa.ca/about-absa/history-of-absa-heritage/heritage/>

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Sánchez, A. N., Raji, D., Rankin, J. L., Richardson, R., Schultz, J., West, S. M., & Whittaker, M. (2019). AI Now 2019 report. [https://ainowinstitute.org/AI\\_Now\\_2019\\_Report.pdf](https://ainowinstitute.org/AI_Now_2019_Report.pdf)

Deon. (n.d.). An ethics checklist for data scientists. <https://deon.drivendata.org>

Gently, D. (2018). Pressure [Photograph]. Flickr. <https://www.flickr.com/photos/6x7/25879810377/>

Hagendorff, T. (2019). The ethics of AI ethics: An evaluation of guidelines. Minds & Machines. doi:10.1007/s11023-020-09517-8

Low, K. (2016). The Human Venture Institute mapbook (16th edition). Action Studies Institute.

O'Neill, C. (2016). Weapons of math destruction. Crown Brooks.

Provost, F., & Fawcett, T. (2014). Data science for business: What you need to know about data mining and data-analytic thinking. O'Reilly Media.

The National Aeronautics and Space Administration. (2020). Pale blue dot revisited [Photograph]. Flickr. <https://www.flickr.com/photos/nasacommons/49533887268/>

Valerie. (2012). Tool usage [Photograph]. Flickr. <https://www.flickr.com/photos/ucumari/7319932060/>

## Thanks.

contact | presentation | ethics resources  
[www.nickkal.com](http://www.nickkal.com)

Selected License  
Attribution-NonCommercial-ShareAlike 4.0 International



Except where otherwise noted, this work is licensed under  
<https://creativecommons.org/licenses/by-nc-sa/4.0/>