



# Philosophy & Ethics

## Module 4: Virtue ethics and Care ethics

Simon Coghlan





## **WARNING**

**This material has been reproduced and communicated to you by or on behalf of the University of Melbourne in accordance with section 113P of the Copyright Act 1968 (Act).**

**The material in this communication may be subject to copyright under the Act.**

**Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.**

**Do not remove this notice**



# Learning outcomes

At the end of this module, you should be able to:

- Describe the ethical theories of virtue ethics and care ethics
- Contrast these theories with utilitarianism and deontology
- Begin to apply these ethical theories to a case study involving AI



Images herein: Unsplash, Wiki

Simon Coghlan



# Ethics theories

1. Utilitarianism – consequence based
2. Deontology – rule based
3. Virtue ethics – character based
4. Ethics of care – relationship based

Principlism – simplified integration of 1-4



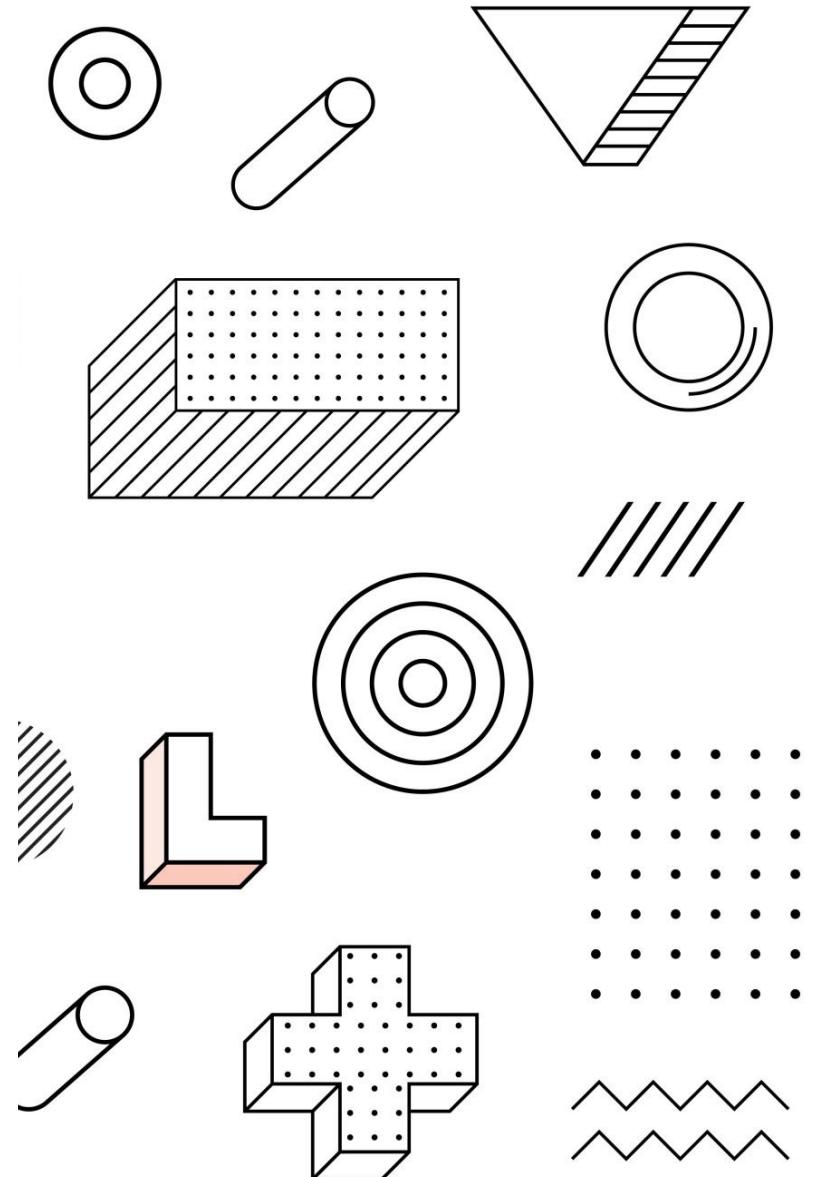
Images herein: Wiki commons and unsplash



# Essay: questions?

**Task 1:** Argue *in favour* of the proposed use of AI (monitoring truck drivers or the smart medical cabinet) using either U or D (not both)

**Task 2:** Argue *against* the proposed use of AI (monitoring truck drivers or the smart medical cabinet) using either VE or CE (not both)



# Utilitarianism

Only consequences  
determine rightness

Principle of utility

Best overall consequences  
(weighing good and bad,  
their magnitude and  
probability)



A:  $5P \times 20U$  pleasure and  
 $2P \times 10$  units pain (-10M)  $= 80M$ . (V).

https://www.britannica.com/topic/utilitarianism-philosophy

Simon Coghlan

B:  $1.5P \times 10U$  pleasure and  $2P \times 15M$  pain (-15M)  $= 70M$



→ 50% x 100% pleasure (V)

## Deontology - rules

100% x 100% pleasure

Don't lie or deceive

Don't steal

Show gratitude

Be just

Respect autonomy

Don't harm

Be beneficent

Respect privacy

道德重要



Can clash – if prima facie: which duty is more compelling? (non-Kantian view)

E.g. white lie vs. major lie; small harm vs. big harm

### THREE LAWS OF ROBOTICS

1. A robot must not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

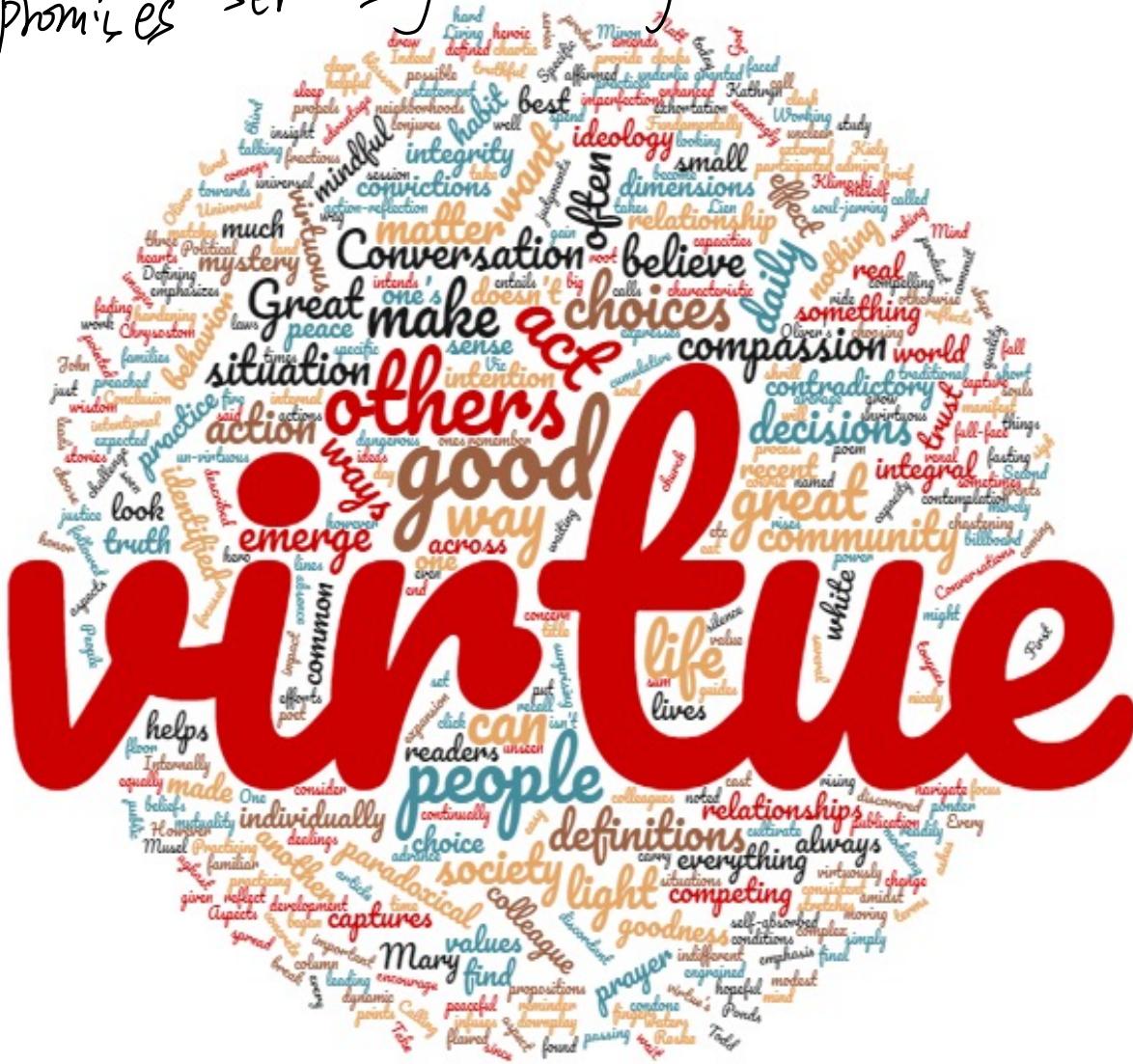
ANDROID AUTHORITY

Simon Coghlan

2 Categorical imperative

universalised liability  
(universalisable) Moral law

1st unconditional duty : apply to everyone  
if everyone breaks promises when it suits them : no one would  
take promises seriously → promise becomes impossible.



2nd: always to respect rational beings not act against/disrespect them.

Never treat them as merely means (deceiving, ...)



# Who are your ‘moral heroes’?



**Siddhattha Gotama:** Perfection of moral character (*sīla pāramitā*)



## *De* (德)

Exemplary person,  
uprightness or right-  
seeing

Simon Coghlan





Simon Coghlan





Simon Coghlan

# Ethical models?





# Virtue Ethics

Ethics of *character*

Duties and consequences too narrow:  
more to ethical life

Elizabeth Anscombe: vestige of religious  
rules (Christian commandments)

Ancient West: Aristotle (384-322 BCE)-  
*Nichomachean Ethics*

Ancient East: Confucius, Buddha (both  
6-5<sup>th</sup> BCE)

Being a *kind of person* -- praiseworthy,  
warrants respect, emulation



道德明陞是諮詢 (ethical review)



Ethics: aims not just at right *action*,  
but good way of *being*

Deontology/consequentialism:  
ignore centrality of  
feeling/emotion/attitude

Aretaic ethics: excellence of  
character (cf. intellect)

Kant: shopkeeper not ripping-off (4/7, 15)  
customers: pure duty/good will vs.  
caring about (2) Ethical  
virtue respect

~~customer~~  
interests of customers





## Responsible Wellbeing

<https://medium.com/age-of-awareness/focusing-on-virtues-and-character-strengths-probably-the-smallest-but-most-decisive-step-towards-c36baaff6325>

LOVE	PERSPECTIVE	CREATIVITY	APRECIATION OF BEAUTY & EXCELLENCE	GRATITUDE	PRUDENCE
CURIOSITY	HUMILITY & MODESTY	SPIRITUALITY	FAIRNESS	LEADERSHIP	HUMOR
PERSISTENCE	OPEN-MINDEDNESS	KINDNESS	CITIZENSHIP	INTEGRITY	FORGIVENESS & MERCY
SELF-REGULATION	HOPE	SOCIAL INTELLIGENCE	LOVE OF LEARNING	BRAVERY	VITALITY



# Virtues & Vices

**Justice**

**Generosity**

**Compassion**

**Courage**

**Temperance**

**Gratitude**

**Honesty**

**Trustworthy**

**Love**

**Glad for others' success**

**Humility**

**Patience**

**Injustice**

**Stingy**

**Callousness/cruelty**

**Cowardice**

**No self-control/greed**

**Ingratitude**

**Mendacity**

**Untrustworthy**

**Hate/Indifference**

**Envy**

**Pride**

**Impatience**



# Aristotle: Virtues (and Vices)

Dispositions or traits

Cultivated over time

Habit – settled dispositions

Trained as children – reward/punish/modeling

Can learn these traits

Can get better (towards perfection) or worse (corrupted)

*Right* feeling, *right* action, *right* way, *right* amount, to *right* people

Kant: 'good will'/follow duty (NOT feeling/inclination)

Aristotle: right attitudes/motivations/emotions

Example: Visit sick friend

Good friend

Kant      duty demands

⑥

Compassion / Virtue of friendship





# Aristotle: Doctrine of Golden Mean

Right feeling/action/thoughts

Character traits: on a spectrum

Virtues: the sweet spot

Excess and deficiency = vice

Affected by context

~~feeling~~/action

Example: lazy student steals your hard work

How should you react and feel? What is the virtue here? The vice(s)?

Indifference---Righteous indignation---Spite

response: deficiency ~~feelings~~  
~~not enough~~  
~~too little~~

excess: extremely outraged

too much

do some bad

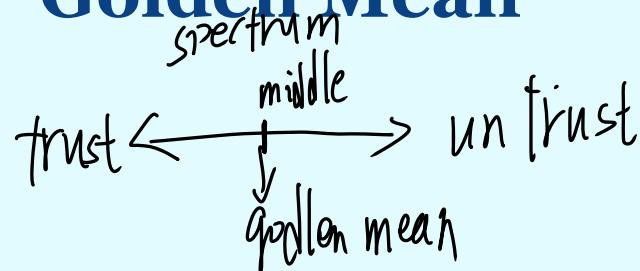
Simon Coghlan

high amount of anger





# Aristotle: Doctrine of Golden Mean



Gullible---Right trust---Cynical

Affected by situation/context e.g. AI in Netflix vs. predictive policing

Also: people might disagree over how to *apply* a virtue e.g. some will say that AI policing can be just, others will say a just person will *never* support it!

You need to use *judgement* and *reasons* to explain what is virtuous and why

e.g. Predictive policing: What are its risks and benefits? History of oppression? What virtues are relevant? Why? Consider opposite points of view?

i.e. virtue ethics is not a simple *formula*. Virtue ethicists may disagree with each other!

# The Virtues

## Vice (Deficiency)

Cowardly  
Addictive  
Stingy  
Self-deprecating  
Boorish  
Quarrelsome  
Melancholy  
Depraved  
Envious  
Mean  
Lazy

## Virtue (Mean)

Brave  
Temperate  
Generous  
Truthful  
Witty  
Friendly  
Spirited  
Conscientious  
Indignant  
Benevolent  
Industrious

## Vice (Excess)

Rash  
Ascetic  
Extravagant  
Boastful  
Buffoonish  
Bootlicking  
Boisterous  
Deferential  
Retributive  
Self-sacrificing  
Single-minded



# Relation to duty

Criticism: VE doesn't tell us what to *do*, only how to *be*

## Phronesis

- Practical wisdom (virtue)
- Apply moral ideas/rules *contextually*
- VE versatile, flexible, sensitive to situations

## What a virtuous person would do?

- How do we know that?
- We first learn from education, training, school, parents
- Aristotle: young people and ethics
- And we begin to look to moral exemplars
- Moral rules follow from this



Rosalind Hursthouse (1943-)



# Flourishing: the good life

Aristotle: virtues are constitutive of good life

Example:

Ben - pretends to be virtuous, but is greedy, callous, unjust, dishonest, unfaithful and has loads of fun

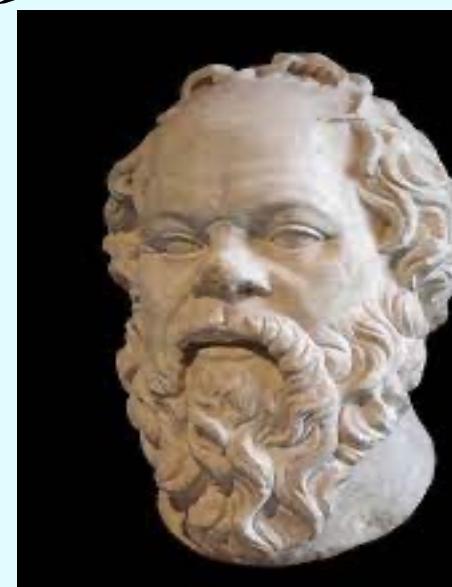
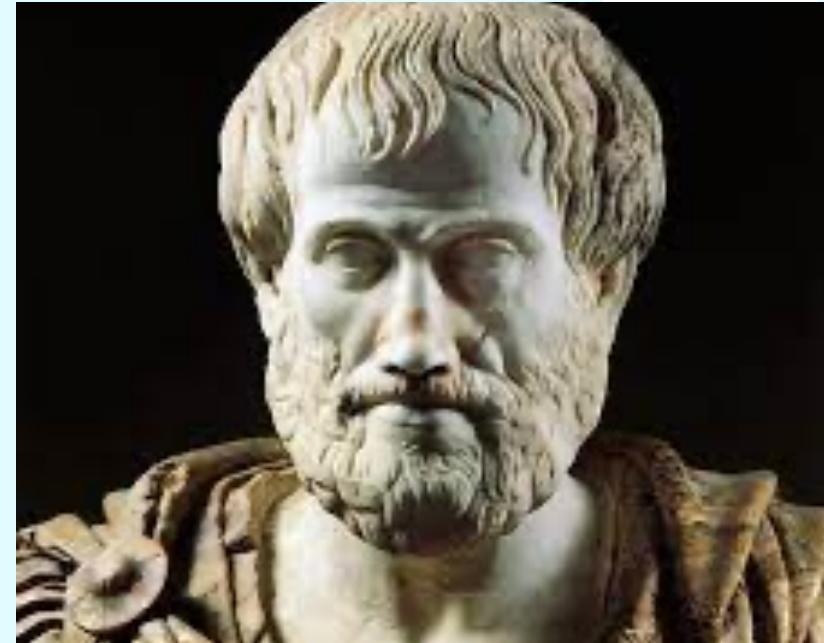
Lei – virtuous and good, but is poor and struggles with chronic illness ↗ from VE Lei has better life

Without virtues: your life goes badly, even if rich, famous, successful

Ethical relationships part of good life

Aristotle: happiness = flourishing = *eudaimonia*

Socrates: 'better to suffer evil than to do it'





# Virtuous exemplars?

Joy Bouamwini/Timnit Gebru

Stood up to Big Tech

Called for fairness, transparency,  
diversity in AI development

What virtues might they have?

Just? Honest? Courageous?

Assessing tech: What would a virtuous  
person do or think about this?





# Shannon Vallor: Techno-virtues

技术结合的  
力量

virtuous 行为

行之有效的

- “technological convergence: discrete technologies merging...magnify[ing] their scope and power”
- Need “context-adaptive moral reasoning”
- Phronesis. *Not* broad rules e.g. Kantian/U
- Technology
  - Growing techno-social opacity
  - Unpredictable consequences
  - Future individuals – human, nonhuman



Need *techno-social wisdom*

Has trans-cultural resonance



# Activity

Will social media

What could a

Vallor: avoid techno-utopianism, dystopianism, determinism

Simon Coghlan

Virtue ethics	Difference between VE & Utilitarianism	Difference between VE & Deontology	Practical wisdom & the 'Golden mean'	Criticisms
An ethical theory that emphasizes that being a <b>virtuous individual</b> is central to making ethical decisions. In VE, one should uphold <b>virtues</b> and avoid <b>vices</b>  What constitutes ethical good is contextual	Utilitarianism looks at the consequences of actions and focuses on maximizing happiness/utility for the largest number of people  VE focuses on individual's moral character	Deontology focuses on individuals doing good by following rules based on moral reason and duty  VE focuses on the character of a person to determine what are 'good' actions	Application of moral ideas/rules in context - asking what a virtuous person would do - offers framework for doing good  The golden mean is the desirable and moral middle ground between two extremes. On the low extreme, we have the vice of deficiency, and on the high extreme, excess.	Virtuous person still acting for reasons - a 'moral' person is still acting on rules, duties, principles, concerns for consequences, utility  Too vague - No specific guidance on moral dilemmas  Duty/consequences more important in Ethics than the individual



# Criticism of VE (from U and D)

- Virtuous people still need to act on/for *reasons*
  - Reasons include rules, duties, principles, concern for consequences/utility
  - (Yes, but the *virtues* help us decide *those things*).
- Too vague
  - No specific guidance on moral dilemmas
  - (Think: What virtues? What exemplars? How to *apply virtues*? What are the circumstances? Consider the details)  

- Duty or *consequences* are *primary*
  - *Character* only valuable instrumentally
  - (*Character* necessary for having good judgement about duties/consequences)  






# Care Ethics

Feminist ethics

Carol Gilligan

Lawrence Kohlberg: Moral development of children

*Heinz' wife: terminally ill*

*A chemist has life-saving drug*

*Overpriced at \$20,000*

*Heinz cannot afford; can't get a loan*

*Should he steal the drug from chemist?*

Simon Coghlan



# Male vs female responses

Jake: Steal drug!

- Human life more basic value than property
- Fairness

(

Amy: Relationships and caring

- Talk to chemist and emphasize dire situation of wife - empathy (care compassion)
- Stealing>jail>wife left alone

Possibly same conclusion, but...

Masculine vs feminine perspective

Masculine: “ethics of justice” and “rights”

Kholberg: Masculine higher – based in detached thinking, rules, principles, consequences





# Carol Gilligan

*In a Different Voice* (1982)

Women's voices sidelined; seen as atypical

Feminine voice – not co-extensive with “gender”

Cf. rationalism, principles, severe impartiality, lack of emotion

Very critical of U and D!

Relational and interpersonal – emotions and actions

Attending, listening, loving, feeling, taking responsibility  
*(Any effort as way of thinking)*

Recognising vulnerability, powerlessness  
*(Vulnerability, Powerlessness)*

Protecting relationships





# Joan Tronto

## Care

*"A species of activity that includes everything we do to maintain, contain, and repair our 'world' so that we can live in it as well as possible. That world includes our bodies, ourselves, and our environment"*

Sustain our own and others' basic needs,  
alleviate pain/suffering, enable mutual  
flourishing



# Care

Care begins in close relations e.g. family

Often face-to-face and embodied

Relationship: carer and cared-for –  
interdependence

Mutual vulnerability: all might suddenly  
need care

Self-sufficiency myth (cf. Kant's autonomous  
person)

Essential – healthy society

Part of VE?

But care central to good character/action

Simon Coghlan





# Joan Tronto

Care involves:

- (1) **attentiveness**: inclination to become aware of need
  - (2) **responsibility**: willingness to respond to need  
*care, puzzle*
  - (3) **competence**: skill providing good successful care  
*(from needs to needs)*
  - (4) **responsiveness**: consideration of the position of others as they see it [empathy], recognition of potential for abuse in care, reviewing and adjusting
- 
- Various emotional responses, attitudes, inclinations (e.g. empathy, compassion)
  - Intelligence and judgement





# Eva Kittay

Emphasises human dependency

Disabled, impaired, disadvantaged people

E.g. Sesha: can't walk, talk, or read

Help to flourish as best she can

E.g. exposure to music human language, etc.

Society should assist - justice





# Criticisms of CE

- Too vague (like VE)

CE: can also use principles, rules that emerge from caring. Some vagueness is important – ethics is contextual

- ~~過份理想化~~  
Romanticises motherhood? Slave morality? Depicts carers (often women) as too self-sacrificing?

CE: Care should be mutual, not women only; men are carers; carers need care too



# Criticisms

- Feminine perspective a stereotype?

CE: Many women stress impartial justice; men who stress care; *not* saying that care is completely gendered



- Parochial to own people/circle? Need impartial justice? ....To strangers, third-world, intersectional oppressions (e.g. trans black people)

CE: True, care often stronger for those we are in special relations with

But: stresses vulnerability and oppression

And: Justice an *extension of caring*

Care can lead to and be involved in justice (e.g. view others as children, parents, brothers, sisters in need of care)

Can have caring relations beyond one's immediate circle





THE UNIVERSITY OF  
MELBOURNE

## CARE ETHICS

### Care ethics

Comes from feminist ethics

Relational and interpersonal

Does not emphasize rules/consequence

Emphasizes care

### Difference between CE & Utilitarianism

Utilitarianism doesn't consider relationships between people, just on maximizing happiness/utility for the largest number of people.

CE focuses on mutual flourishing by taking care of other's basic needs and alleviating pain/suffering

### Difference between CE & Deontology

Deontology focuses on individuals following rules based on moral reason and duty. Treats each person as entirely autonomous self-sufficient.

CE acknowledges that we are all interrelated, vulnerable to eachother and should care for others

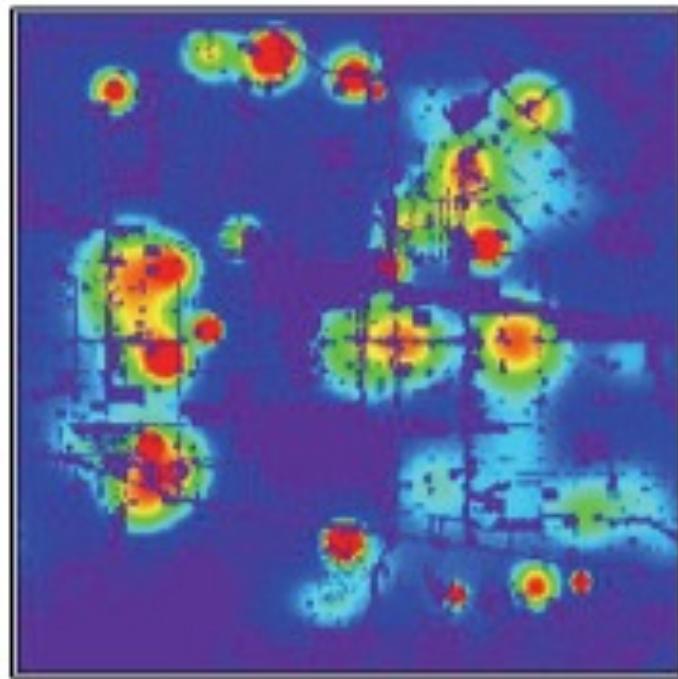
### Criticism

Too vague - no rules/principles to work from

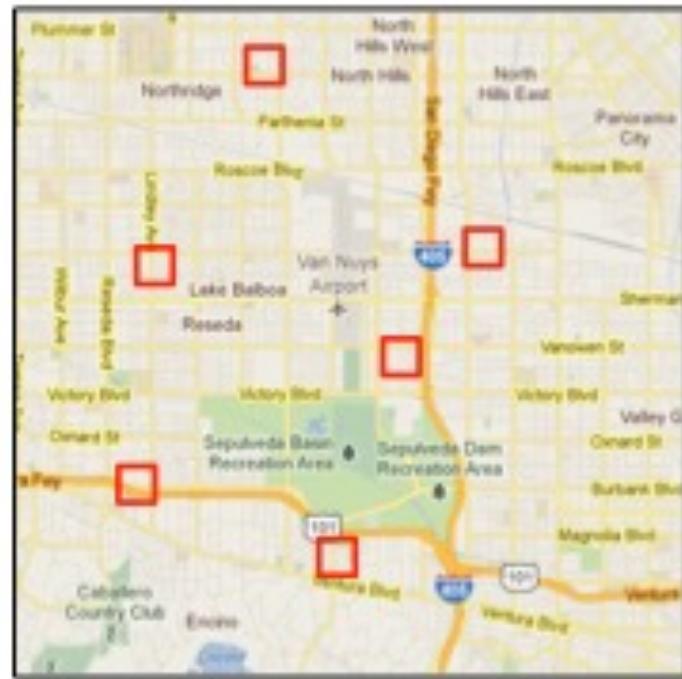
- Rules about morality emerge from caring
- Vagueness needed as all ethics are contextual

Romanticizes & stereotypes feminine care-men should be carers too

# Predpol



**TACTICAL AMBIGUITY**  
*rear-view mirror heat map*



**TACTICAL CLARITY**  
*forward-looking PredPol boxes*

# Principlism

- Distill theories into handy principles for AI ethics
- Midlevel principles: b/w theory and detailed rules
- Theory can *guide* midlevel principles

1. Data (数据要去训练. → 数据可能犯隐私)  
 2. 什么样的罪行算 minor. offences. 什么样的罪行算 major?  
 vagrant (流浪), 必须会加重警察的负担 (负担)  
 Small quantity of crimes (没有造成一些潜在的危害)  
 Consuming kids jump subway turnstalls

⇒ 警察的思维.  
 ⇒ 要 arrest more people.  
 扩大 Patrol range.

# Principlism: 4 principles +1

Derived from medical ethics

## 1. Non-maleficence – do no harm

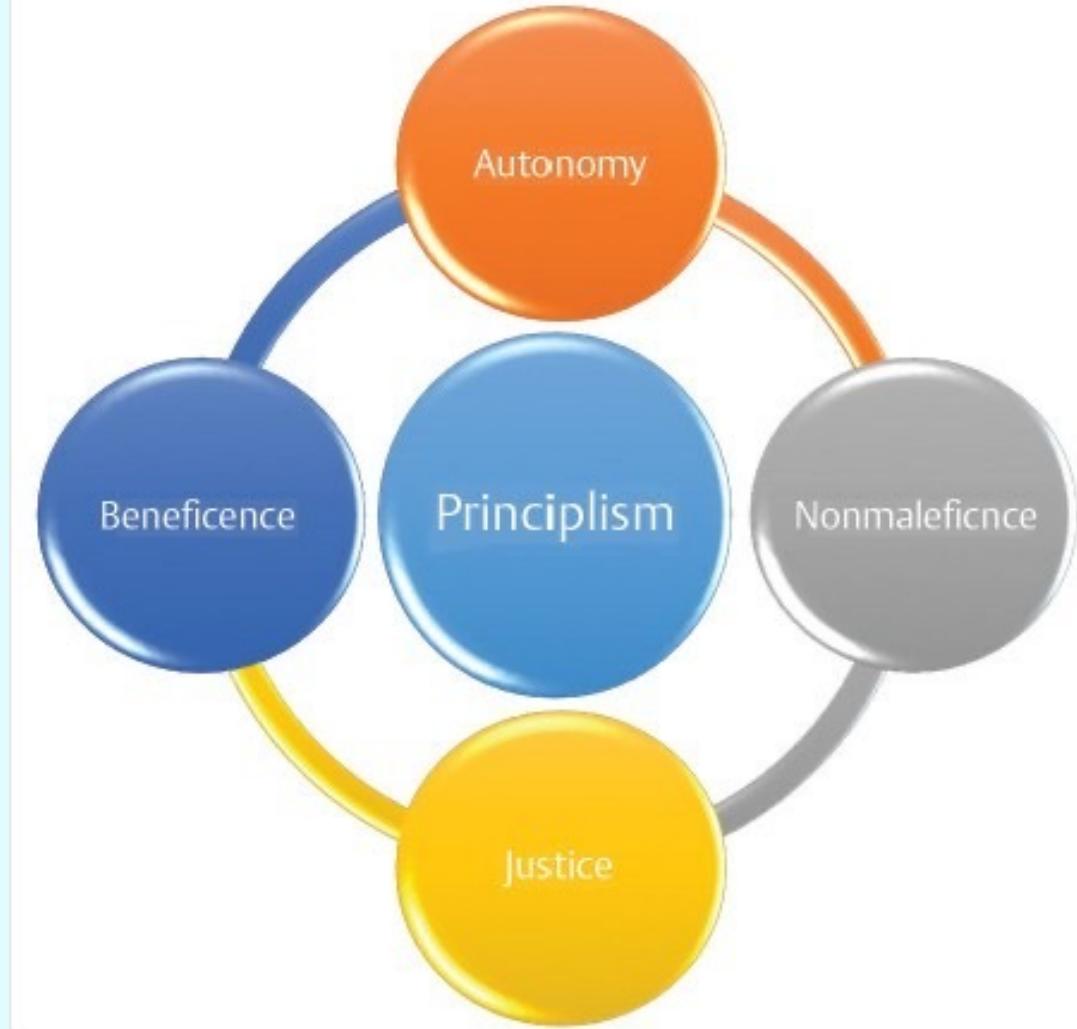
- Predict harm, avoid causing harm, minimize harm, short and long term

## 2. Beneficence – do good

- Anticipate good outcomes, short and long term

## 3. Respect autonomy – respect people's values, choices, life plans

- Understand what others' value, don't override their choices, be honest,





# Principlism

## 4. Justice – fairness

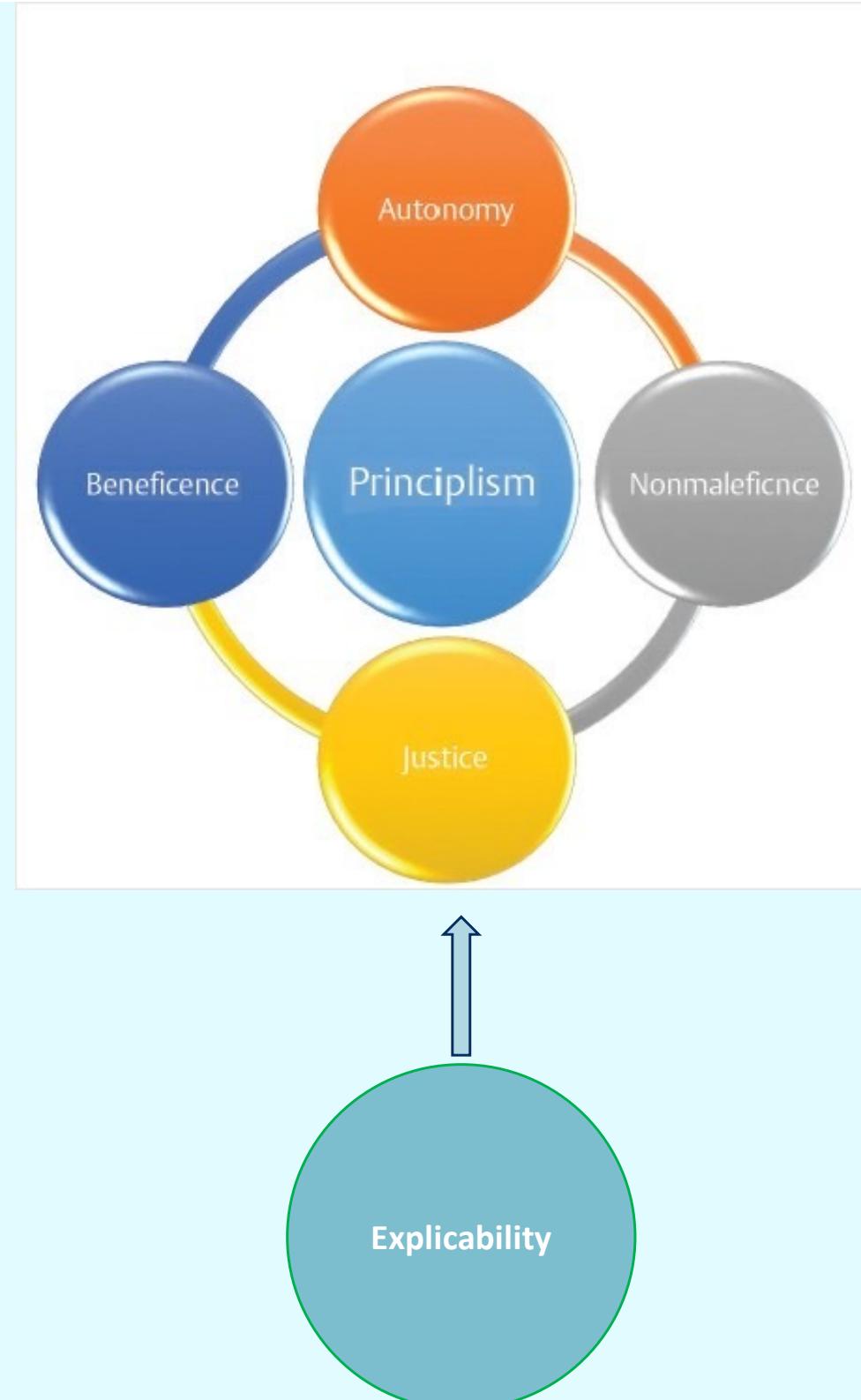
- Distribute benefits and harms fairly, fair processes, don't unfairly discriminate

## 4+1. Explicability – transparency and accountability (Floridi\*)

- Complements the 4 principles
- Ensure those potentially impacted have sufficient understanding of the AI and that relevant people are held to account

Principles: need to be balanced against one another; all are 'equal'

\*Floridi, Luciano, et al. "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations." *Minds and Machines* 28.4 (2018): 689-707.





Simon Coghlan

GRAZIE VINAKA 합니다

TERIMA KASIH 다

THANK YOU TAKK

謝 謝 ありがとう merci