

Föreläsning på MVE420: Nya teknologier, global risk och
mänsklighetens framtid

Etik med relevans för globala frågor

30 mars 2021

Olle Häggström

<https://www.chalmers.se/en/Staff/Pages/olle-haggstrom.aspx>

<http://haggstrom.blogspot.com/>

Ur kurs-PM:s instruktioner för era projekt:

Ur kurs-PM:s instruktioner för era projekt:

“Rapporten bör också ta upp hur vi (mänskligheten) kan agera för att ge bästa chanser till ett positivt utfall, och vilka moralfilosofiska idéer som (utifrån analysen till Deadline P2) ligger till grund för dessa överväganden. Detta kan betyda att lyfta några relevanta värderingsfrågor som aktualiseras av de nya teknologierna.”

Men vad har etik och moralfilosofi med nya teknologier att göra?

Men vad har etik och moralfilosofi med nya teknologier att göra?

Jag börjar med ett konkret exempel.

Men vad har etik och moralfilosofi med nya teknologier att göra?
Jag börjar med ett konkret exempel.

[Home](#) [Who We Are](#) [Activities](#) [Existential Risk](#) [Get Involved](#) [Contact](#)

 *Technology is giving life
the potential to flourish
like never before...* 

[News:](#) [AI](#) [Biotech](#) [Nuclear](#) [Climate](#) [Partner Orgs](#)

*This open letter was announced July 28 at the opening of the IJCAI 2015 conference on July 28.
Journalists who wish to see the press release may contact [Toby Walsh](#).
Hosting, signature verification and list management are supported by FLI; for administrative questions about this letter,
please contact [Max Tegmark](#).*

AUTONOMOUS WEAPONS: AN OPEN LETTER FROM AI & ROBOTICS RESEARCHERS

Autonomous weapons select and engage targets without human intervention. They might include, for example, armed quadcopters that can search for and eliminate people meeting certain pre-defined criteria, but do not include cruise

Ur brevet:

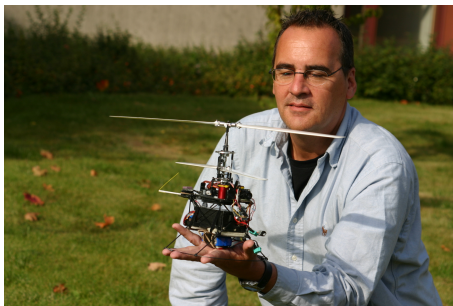
Ur brevet: “If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow.

Ur brevet: “If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce.

Ur brevet: “If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. It will only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc.

Ur brevet: “If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. It will only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and selectively killing a particular ethnic group. We therefore believe

Ur brevet: “If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. It will only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and selectively killing a particular ethnic group. We therefore believe that a military AI arms race would not be beneficial for humanity.”



Patrick Doherty, professor i datavetenskap vid Linköpings universitet, augusti 2015:



Patrick Doherty, professor i datavetenskap vid Linköpings universitet, augusti 2015:

“Jag har inte skrivit under det öppna brevet, ty det finns inga onda eller goda teknologier, det finns bara onda och goda tillämpningar.”

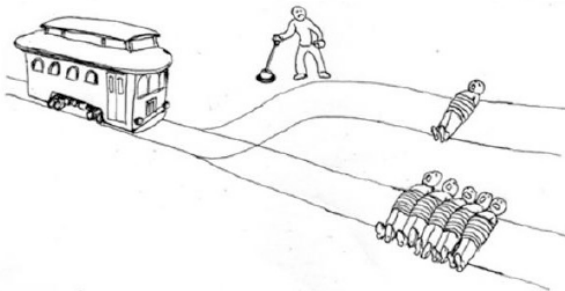
Patrick Doherty och jag tycker olika i denna etikfråga.

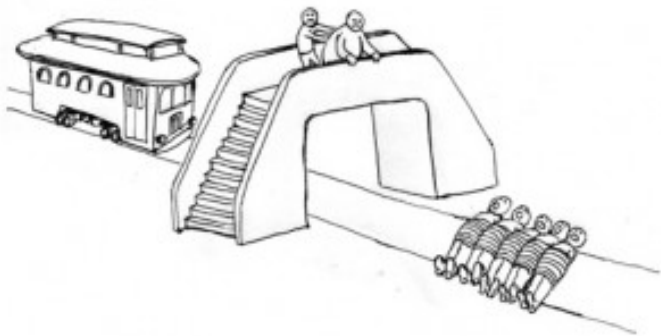
Patrick Doherty och jag tycker olika i denna etikfråga.

Men det hela behöver inte stanna vid “han tycker så och jag tycker så”. Det akademiska ämnet *moralfilosofi* (även kallat *etik*) finns till för att ge struktur åt etiska diskussioner och möjliggöra rationellt grundade resonemang kring dem, istället för att förfalla till allmänt tyckande.

Populärt moralfilosofiskt diskussionsämne på senare år:
trolleyologi.

Populärt moralfilosofiskt diskussionsämne på senare år:
trolleyologi.







Karim Jebari @KarimJebari · 4h

När man undervisar i filosofi är det alltid någon student som säger något i stil med "men om jag sliter loss spaken och slänger ned den på spåret kanske jag kan stoppa tåget". Det är väldigt tacksamt, för då kan du förklara att det är att missa poängen med ett tankeexperiment.

[Translate Tweet](#)



1



1



Karim Jebari @KarimJebari · 4h

Syftet med experimentet är inte att diskutera de bästa sätten att rädda folk från tåg, utan att fundera på de etiska intuitionerna som får oss att tveka (eller inte).

[Translate Tweet](#)



1



1



Karim Jebari @KarimJebari · 4h

Med andra ord: tankeexperiment är teoretiska redskap för att resonera kring abstrakta principer. De är inte scenarion för att planera inför en viss situation eller träna för att lösa praktiska problem.

[Translate Tweet](#)



1



2



Karim Jebari @KarimJebari · 4h

Allt detta är enormt okontroversiellt. Det som dock är märkligt för mig, är att så fort filosofer (och andra) ska börja ägna sig åt tillämpad etik, så glömmes de distinktionen mellan ett tankeexperiment och ett scenario.

[Translate Tweet](#)



1

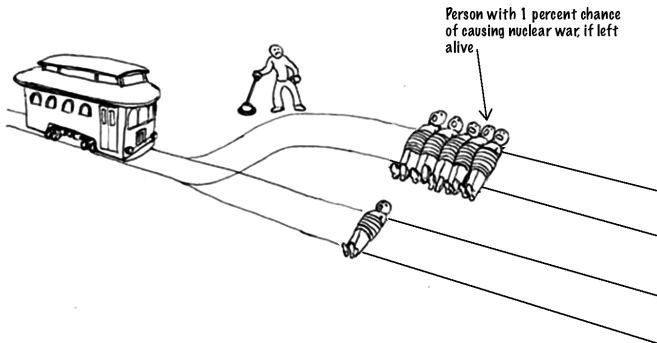


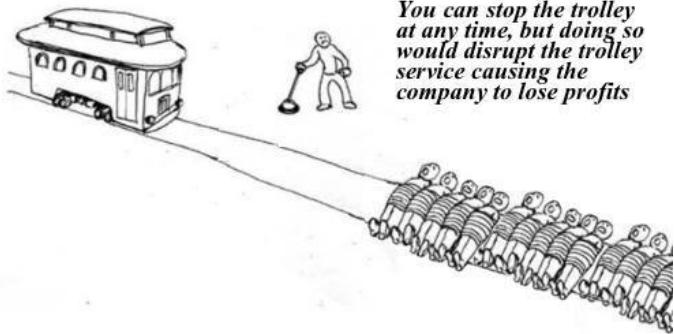
5



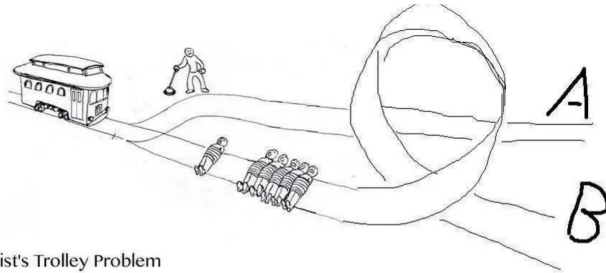
Karim Jebari @KarimJebari · 4h

Detta är särskilt påtagligt i diskussionen om självkörande bilar. Det har uppstått en hel litteratur som handlar om att postulera att trolley-problemet och andra tankeexperiment faktiskt handlar om verkligheten.





*You can stop the trolley
at any time, but doing so
would disrupt the trolley
service causing the
company to lose profits*



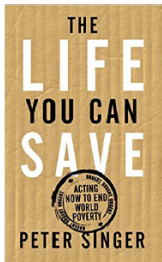
Hedonist's Trolley Problem

The track is heading towards B.

If you pull the lever, it will switch to A but it won't do the totally sick loop-da-loop.

Ett annat berömt tankeexperiment:

Ett annat berömt tankeexperiment:



Tre olika moralfilosofiska ramverk:

Tre olika moralfilosofiska ramverk:

1. Konsekventialism: en handling bedöms utifrån dess *konsekvenser*.

Tre olika moralfilosofiska ramverk:

1. Konsekventialism: en handling bedöms utifrån dess *konsekvenser*.

2. Deontologisk etik: en handling bedöms utifrån dess *inneboende värde*.

Tre olika moralfilosofiska ramverk:

1. Konsekventialism: en handling bedöms utifrån dess *konsekvenser*.

2. Deontologisk etik: en handling bedöms utifrån dess *inneboende värde*.

3. Dygdetik: en handling bedöms utifrån hur väl den överensstämmer med *vad en dygdig person skulle ha gjort*.

Tre olika moralfilosofiska ramverk:

1. Konsekventialism: en handling bedöms utifrån dess *konsekvenser*.

1a. Utilitarism förordar de handlingar som leder till mest *nytta* i världen, där nytta kan likställas med t.ex. välbefinnande (hedonisk utilitarism) eller uppfyllda preferenser (preferensutilitarism).

2. Deontologisk etik: en handling bedöms utifrån dess *inneboende värde*.

3. Dygdetik: en handling bedöms utifrån hur väl den överensstämmer med *vad en dygdig person skulle ha gjort*.

Tre olika moralfilosofiska ramverk:

1. Konsekventialism: en handling bedöms utifrån dess *konsekvenser*.

1a. Utilitarism förordar de handlingar som leder till mest *nytta* i världen, där nytta kan likställas med t.ex. välbefinnande (hedonisk utilitarism) eller uppfyllda preferenser (preferensutilitarism).

2. Deontologisk etik: en handling bedöms utifrån dess *inneboende värde*.

2a. Pliktetik.

2b. Rättighetsetik.

3. Dygdetik: en handling bedöms utifrån hur väl den överensstämmer med *vad en dygdig person skulle ha gjort*.

En annan uppdelning av utilitaristiska moralteorier:

En annan uppdelning av utilitaristiska moralteorier: oavsett val av nyttofunktion kan vi skilja mellan

- ▶ **handlingsutilitarism,**

En annan uppdelning av utilitaristiska moralteorier: oavsett val av nyttofunktion kan vi skilja mellan

- ▶ **handlingsutilitarism**, och
- ▶ **regelutilitarism** (som går ut på att låta samhället styras av *de regler som gör mest nytta*).

Oavsett val av etisk teori gör vi klokt i att skilja mellan **faktapåståenden** och **värdepåståenden**.

Oavsett val av etisk teori gör vi klokt i att skilja mellan **faktapåståenden** och **värdepåståenden**.

Humes lag: Det går inte att härleda värdepåståenden från enbart faktapåståenden.

Oavsett val av etisk teori gör vi klokt i att skilja mellan **faktapåståenden** och **värdepåståenden**.

Humes lag: Det går inte att härleda värdepåståenden från enbart faktapåståenden.



Exempel:

Exempel:

Premiss: Arsenik är ett dödligt gift.

Exempel:

Premiss: Arsenik är ett dödligt gift.

Slutsats: Det är fel att lägga arseknik i sina middagsgästers mat.

Exempel:

Premiss: Arsenik är ett dödligt gift.

Slutsats: Det är fel att lägga arseknik i sina middagsgästers mat.

Detta bryter mot Humes lag och är därför en felaktig slutledning.

Nytt försök:

Nytt försök:

Premiss 1: Arsenik är ett dödligt gift.

Nytt försök:

Premiss 1: Arsenik är ett dödligt gift.

Premiss 2: Du skall icke dräpa.

Nytt försök:

Premiss 1: Arsenik är ett dödligt gift.

Premiss 2: Du skall icke dräpa.

Slutsats: Det är fel att lägga arseknik i sina middagsgästers mat.





Christian Azar svarar, i DN den 21 april 2017:





Christian Azar svarar, i DN den 21 april 2017:

“Det påstås att klimatfrågan blivit så viktig att vetenskapen bör gå före politiken. Det vore ett misstag.



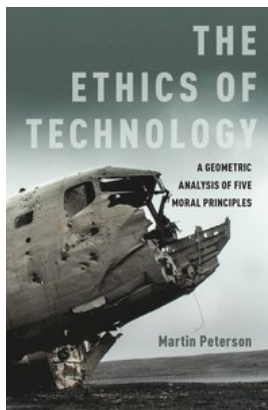
Christian Azar svarar, i DN den 21 april 2017:

“Det påstås att klimatfrågan blivit så viktig att vetenskapen bör gå före politiken. Det vore ett misstag.

[...]

Vetenskaplig kunskap måste naturligtvis spela en mycket viktig roll när vi utformar vår politik inom ett så komplext område som klimatet. Men om vi skulle överlämna klimatfrågan från politiken till vetenskapen, och låta forskare bestämma mål för klimatpolitiken, ja då skulle vi göra våld på vad som är vetenskapens uppgift och vi skulle kraftigt inskränka demokratin.”

De moralfilosofiska ramverk som diskuterats ovan (konsekventialism, deontologisk etik, dygdetik, ...) har ibland en tendens att bli lite för abstrakta för att ge konkret ledning i verkliga tillämpningar. Lite konkretare är de "fem moralprinciper" som filosofen Martin Peterson tar upp i sin bok **The Ethics of Technology**.



1. The Cost-Benefit Principle

- 1. The Cost-Benefit Principle**
- 2. The Precautionary Principle**

- 1. The Cost-Benefit Principle**
- 2. The Precautionary Principle**
- 3. The Sustainability Principle**

- 1. The Cost-Benefit Principle**
- 2. The Precautionary Principle**
- 3. The Sustainability Principle**
- 4. The Autonomy Principle**

- 1. The Cost-Benefit Principle**
- 2. The Precautionary Principle**
- 3. The Sustainability Principle**
- 4. The Autonomy Principle**
- 5. The Fairness Principle**

1. The Cost-Benefit Principle

A technological intervention to which the Cost-Benefit Principle applies is morally right only if the net surplus of the benefits over costs for all those affected is at least as large as that of every alternative.

2. The Precautionary Principle

3. The Sustainability Principle

4. The Autonomy Principle

5. The Fairness Principle

1. The Cost-Benefit Principle

2. The Precautionary Principle

A technological intervention to which the Precautionary Principle applies is morally right only if reasonable precautionary measures are taken to safeguard against uncertain but nonnegligible threats.

3. The Sustainability Principle

4. The Autonomy Principle

5. The Fairness Principle

- 1. The Cost-Benefit Principle**
- 2. The Precautionary Principle**
- 3. The Sustainability Principle**

A technological intervention to which the Sustainability Principle applies is morally right only if it does not lead to any significant long-term depletion of natural, social or economic resources.

- 4. The Autonomy Principle**
- 5. The Fairness Principle**

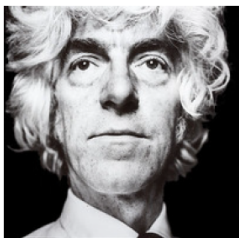
- 1. The Cost-Benefit Principle**
- 2. The Precautionary Principle**
- 3. The Sustainability Principle**
- 4. The Autonomy Principle**

A technological intervention to which the Autonomy Principle applies is morally right only if it does not reduce the independence, self-governance or freedom of the people affected by it.

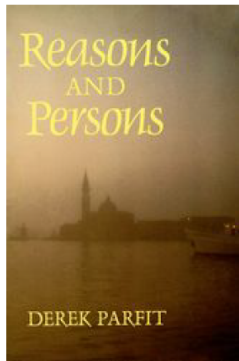
- 5. The Fairness Principle**

- 1. The Cost-Benefit Principle**
- 2. The Precautionary Principle**
- 3. The Sustainability Principle**
- 4. The Autonomy Principle**
- 5. The Fairness Principle**

A technological intervention to which the Fairness Principle applies is morally right only if it does not lead to unfair inequalities among the people affected by it.



Derek Parfit (1942-2017)



Parfit, 1984: “I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes:

Parfit, 1984: “I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes:

(1) Peace.

Parfit, 1984: “I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes:

(1) Peace.

(2) A nuclear war that kills 99% of the world's existing population.

Parfit, 1984: “I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes:

- (1) Peace.
- (2) A nuclear war that kills 99% of the world's existing population.
- (3) A nuclear war that kills 100%.

Parfit, 1984: “I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes:

(1) Peace.

(2) A nuclear war that kills 99% of the world's existing population.

(3) A nuclear war that kills 100%.

(2) would be worse than (1), and (3) would be worse than (2). Which is the greater of these two differences?

Parfit, 1984: "I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes:

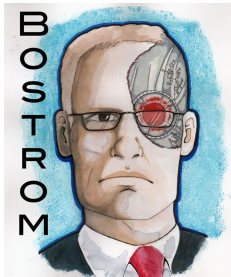
(1) Peace.

(2) A nuclear war that kills 99% of the world's existing population.

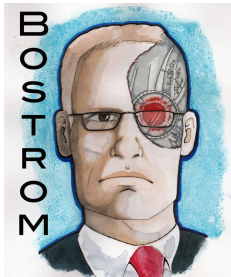
(3) A nuclear war that kills 100%.

(2) would be worse than (1), and (3) would be worse than (2). Which is the greater of these two differences?

Most people believe that the greater difference is between (1) and (2). I believe that the difference between (2) and (3) is very much greater."

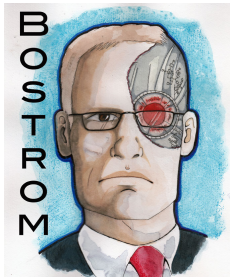


Mycket står på spel. Bostrom (2013) gör några överslagskalkyler. Om vi spelar våra kort rätt, och...



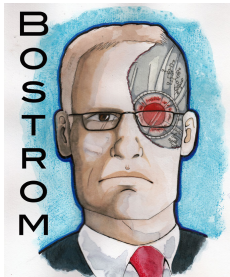
Mycket står på spel. Bostrom (2013) gör några överslagskalkyler. Om vi spelar våra kort rätt, och...

- mänskligheten överlever i 10^9 år,



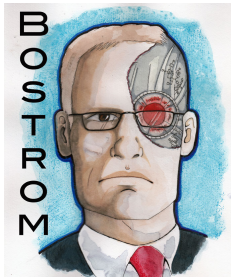
Mycket står på spel. Bostrom (2013) gör några överslagskalkyler. Om vi spelar våra kort rätt, och...

- ▶ mänskligheten överlever i 10^9 år,
- ▶ en befolkning på 10^9 kan upprätthållas,



Mycket står på spel. Bostrom (2013) gör några överslagskalkyler. Om vi spelar våra kort rätt, och...

- ▶ mänskligheten överlever i 10^9 år,
- ▶ en befolkning på 10^9 kan upprätthållas,
- ▶ medellivslängden är 10^2 år,

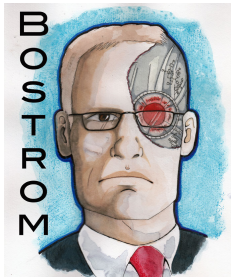


Mycket står på spel. Bostrom (2013) gör några överslagskalkyler. Om vi spelar våra kort rätt, och...

- ▶ mänskligheten överlever i 10^9 år,
- ▶ en befolkning på 10^9 kan upprätthållas,
- ▶ medellivslängden är 10^2 år,

så blir antalet framtida människoliv

$$10^9 10^9 10^{-2} = 10^{16}$$



Mycket står på spel. Bostrom (2013) gör några överslagskalkyler. Om vi spelar våra kort rätt, och...

- ▶ mänskligheten överlever i 10^9 år,
- ▶ en befolkning på 10^9 kan upprätthållas,
- ▶ medellivslängden är 10^2 år,

så blir antalet framtida människoliv

$$10^9 10^9 10^{-2} = 10^{16}$$

Bostrom föreslår även betydligt högre tal baserade på interstellär rymdkolonisering (10^{34}) och/eller uppladdning (10^{54}).



Mycket står på spel. Bostrom (2013) gör några överslagskalkyler. Om vi spelar våra kort rätt, och...

- ▶ mänskligheten överlever i 10^9 år,
- ▶ en befolkning på 10^9 kan upprätthållas,
- ▶ medellivslängden är 10^2 år,

så blir antalet framtida människoliv

$$10^9 10^9 10^{-2} = 10^{16}$$

Bostrom föreslår även betydligt högre tal baserade på interstellär rymdkolonisering (10^{34}) och/eller uppladdning (10^{54}).

Inte alla köper dock den Parfit-Bostromska synen på hur vi skall värdera mänsklighetens undergång.

Inte alla köper dock den Parfit-Bostromska synen på hur vi skall värdera mänsklighetens undergång.



Inte alla köper dock den Parfit-Bostromska synen på hur vi skall värdera mänsklighetens undergång.



Karin Kuhleemann hör till dem som försvarar en så kallad *person-affecting view*, som hävdar att om ett utfall skall räknas som dåligt så måste det vara dåligt *för någon*. Om vi ställer till det så att mänskligheten går under inom kort, så är det dåligt för de 7.8 miljarder människor som lever nu, men inte för Bostroms 10^{16} hypotetiska framtida människor.

Till **Deadline P2** (torsdag 15 april) behöver ni inkomma med en kort (en halv till en sida) preliminär idé om hur de etiska aspekterna på det valda ämnesområdet för ert projekt skall hanteras, strukturerat (i huvudsak) enligt något av följande två scheman.

Praktisk metod för etisk analys: Hanssons förslag

Praktisk metod för etisk analys: Hanssons förslag

1. Identifiera intressenter och deras roller, och beskriv situationen ur deras synvinklar.

Praktisk metod för etisk analys: Hanssons förslag

1. Identifiera intressenter och deras roller, och beskriv situationen ur deras synvinklar.
2. Undersök vilka faktiska förhållanden som påverkar den etiska bedömningen, samt precisera vilken information som saknas.

Praktisk metod för etisk analys: Hanssons förslag

1. Identifiera intressenter och deras roller, och beskriv situationen ur deras synvinklar.
2. Undersök vilka faktiska förhållanden som påverkar den etiska bedömningen, samt precisera vilken information som saknas.
3. Jämför med hur liknande fall skulle ha bedömts.

Praktisk metod för etisk analys: Hanssons förslag

1. Identifiera intressenter och deras roller, och beskriv situationen ur deras synvinklar.
2. Undersök vilka faktiska förhållanden som påverkar den etiska bedömningen, samt precisera vilken information som saknas.
3. Jämför med hur liknande fall skulle ha bedömts.
4. Undersök problemet med flera olika moralteoriers hjälp. Leder de olika teorierna till samma rekommendation?

Praktisk metod för etisk analys: Hanssons förslag

1. Identifiera intressenter och deras roller, och beskriv situationen ur deras synvinklar.
2. Undersök vilka faktiska förhållanden som påverkar den etiska bedömningen, samt precisera vilken information som saknas.
3. Jämför med hur liknande fall skulle ha bedömts.
4. Undersök problemet med flera olika moralteoriers hjälp. Leder de olika teorierna till samma rekommendation?
5. Försök konstruera nya handlingsalternativ.

Praktisk metod för etisk analys: Hanssons förslag

1. Identifiera intressenter och deras roller, och beskriv situationen ur deras synvinklar.
2. Undersök vilka faktiska förhållanden som påverkar den etiska bedömningen, samt precisera vilken information som saknas.
3. Jämför med hur liknande fall skulle ha bedömts.
4. Undersök problemet med flera olika moralteoriers hjälp. Leder de olika teorierna till samma rekommendation?
5. Försök konstruera nya handlingsalternativ.
6. Undersök om föreslagna lösningar är praktiskt genomförbara.

Praktisk metod för etisk analys: van de Poel och Royakkers förslag

Praktisk metod för etisk analys: van de Poel och Royakkers förslag

1. Etisk problemformulering. Vilka värden kan komma i konflikt?
2. Problemanalys, med beaktande av (a) vilka intressenter som finns, (b) relevanta värden, och (c) relevanta fakta.
3. Identifiera handlingsalternativ.
4. Värdera de olika handlingsalternativen på basis av olika moralteorier.
5. Reflektera (*“different ethical frameworks [...] do not necessarily lead to the same conclusion [...] intuitive judgement and common sense are also important”*) och gå tillbaka till steg 3 om det behövs.