

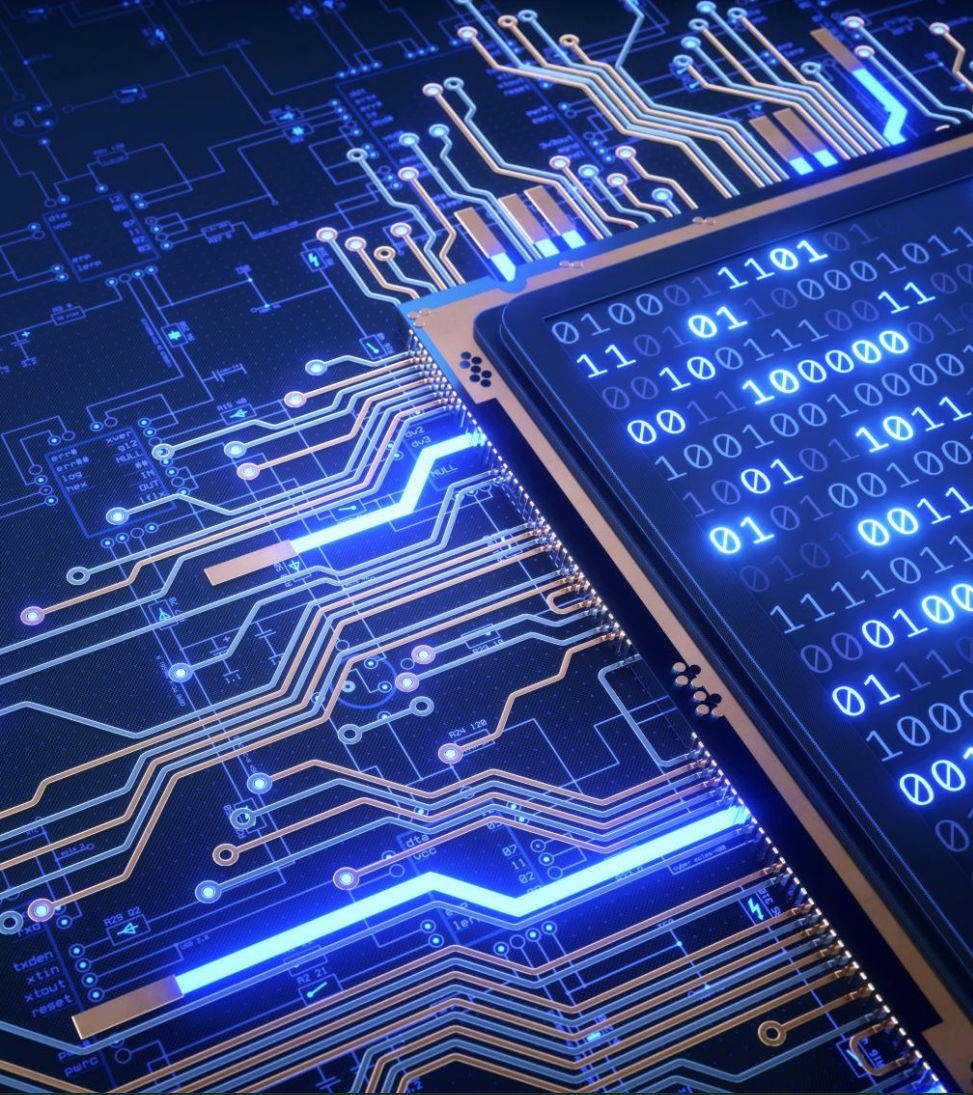
ETIKA UMĚLÉ INTELIGENCE

David Černý

Ústav informatiky AV ČR

LF MU Brno





Struktura kurzu

Základy etiky

Umělá inteligence a její podoby

Obecná umělá inteligence a superinteligence

Existenční rizika

Vybrané systémy AI

LITERATURA

Max Tegmark

ŽIVOT 3.0

Člověk v éře umělé inteligence



LITERATURA

Stuart Russell

JAKO ČLOVĚK

Umělá inteligence a problém jejího ovládnání



argo / dokořán

LITERATURA

NEW YORK TIMES BESTSELLER - 2014

NICK BOSTROM



SUPER
INTELI
GENCE

AŽ BUDOU STROJE GHYTRĚJŠÍ NEŽ LIDÉ

LITERATURA

CROSSOVER → ARGO

NAD PROPASTÍ

*Existenční riziko
a budoucnost lidstva*

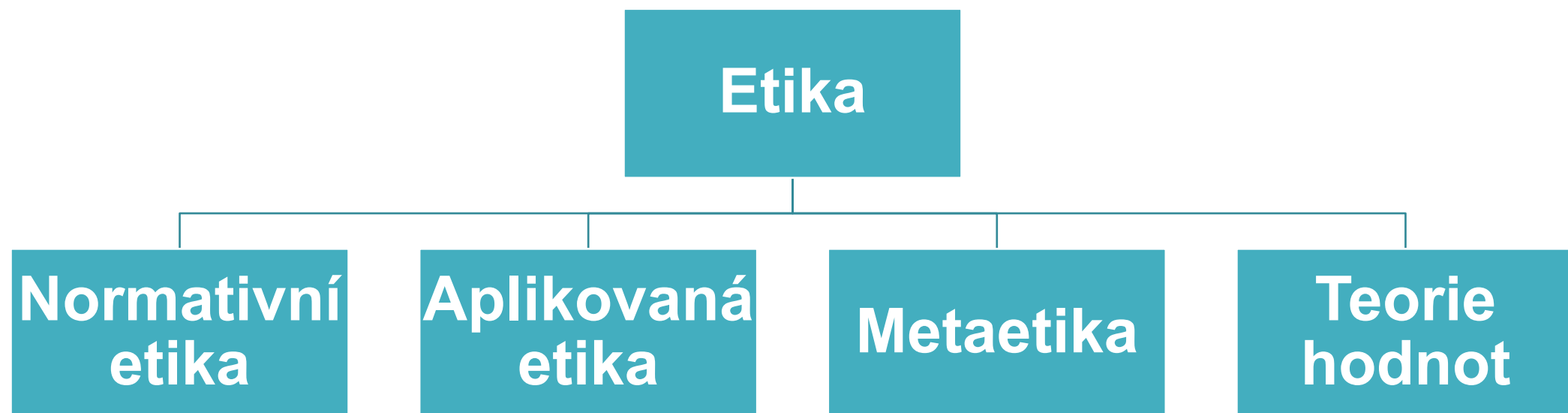
12

Toby
Ord

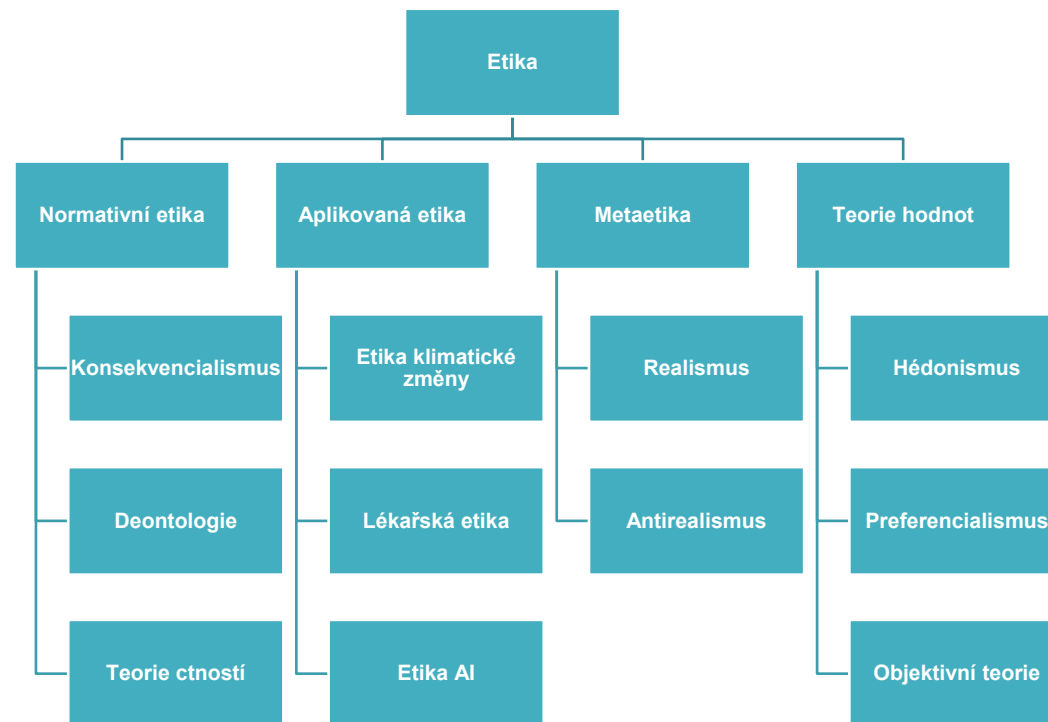
Pokud vše půjde dobře, historie lidstva je teprve na svém počátku. Náš druh pěstuje miliardy let a máde rockvýt do podoby, kterou si dnes ještě neumíme představit. Tato budoucnost je však nyní ohrožena. Kniha je založena na mnohaletém výzkumu existenčních rizik, jimž jako lidstvo čelíme, od klimatické změny přes nekontrolovatelné šíření patogenů až po vývoj AI. Autor tato rizika pojímá jako součást velkého příběhu lidstva a je přesvědčen, že vytvářet nejpravděpodobnější scénáře průběhu hrozících katastrof i toho, jak tato rizika minimalizovat, je největším morálním úkolem současnosti.

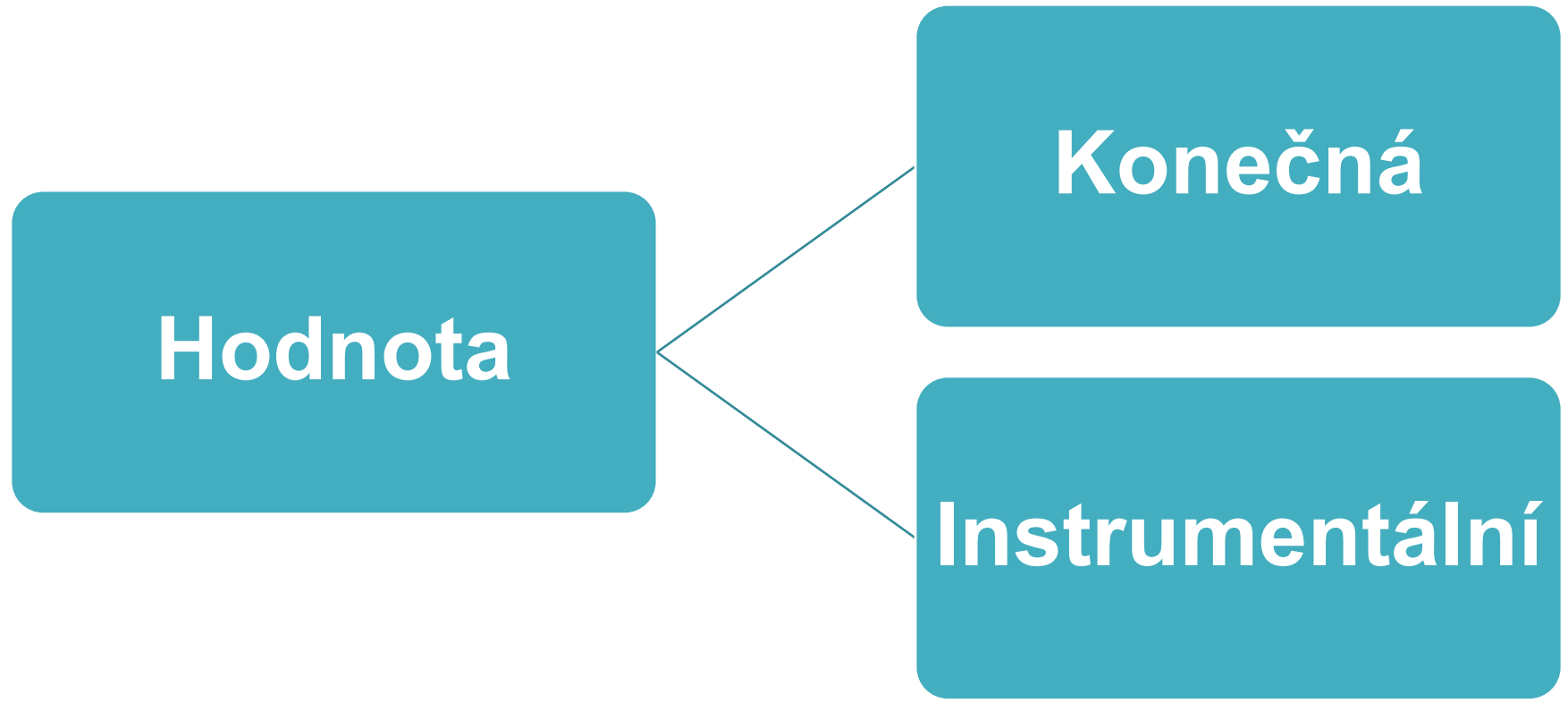


Etika



Etika





Hédonismus

Hédonismus

**Konečné
hodnoty:**

**Pozitivní:
potěšení**

**Negativní:
bolest**



HAPPINESS

MACHINE

STROJ NA ZKUŠENOSTI



Splněné touhy

```
graph TD; A[Hodnoty] --> B[Pozitivní: splněné touhy]; A --> C[Negativní: nesplněné touhy];
```

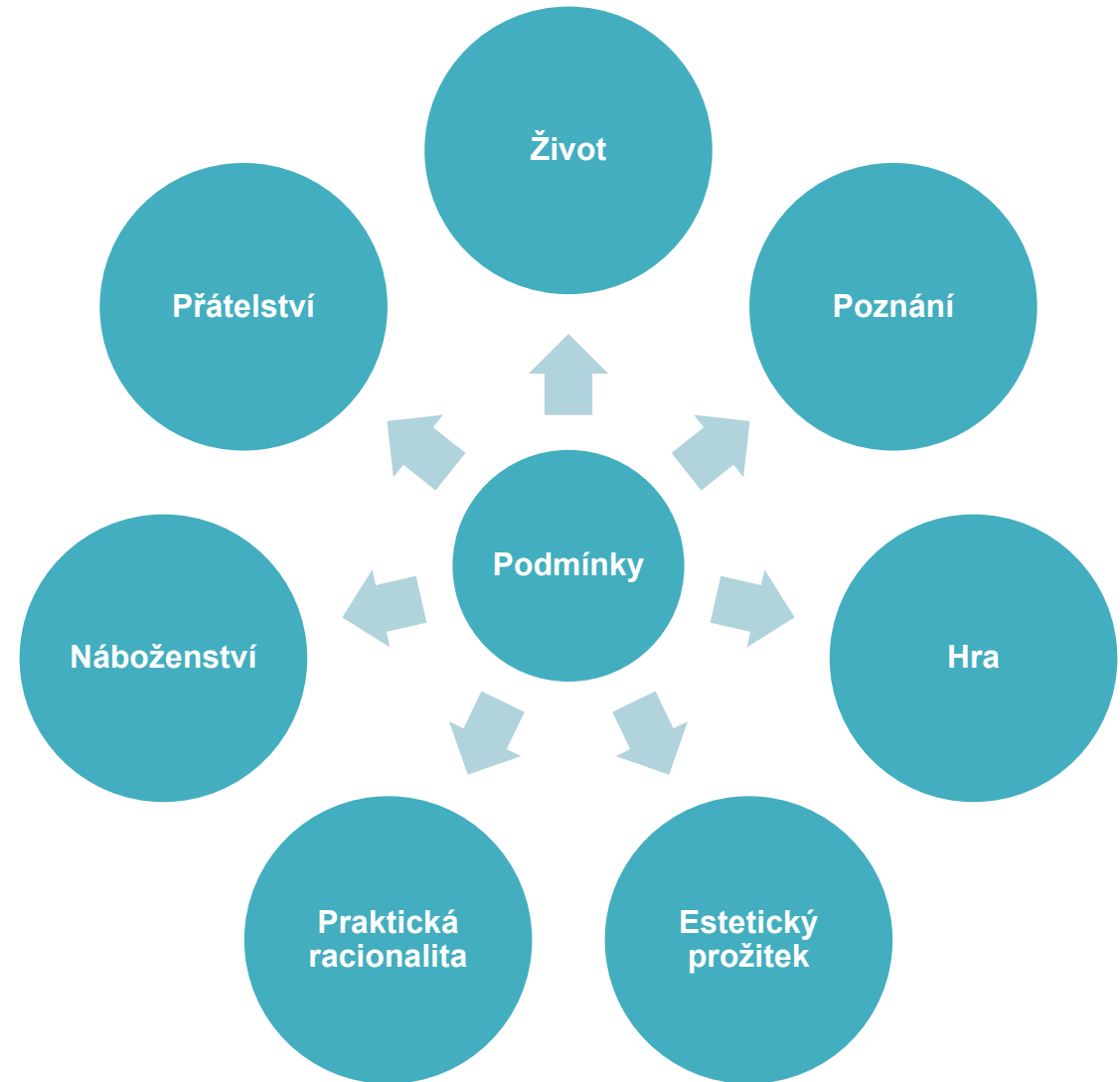
Hodnoty

Pozitivní:
splněné
touhy

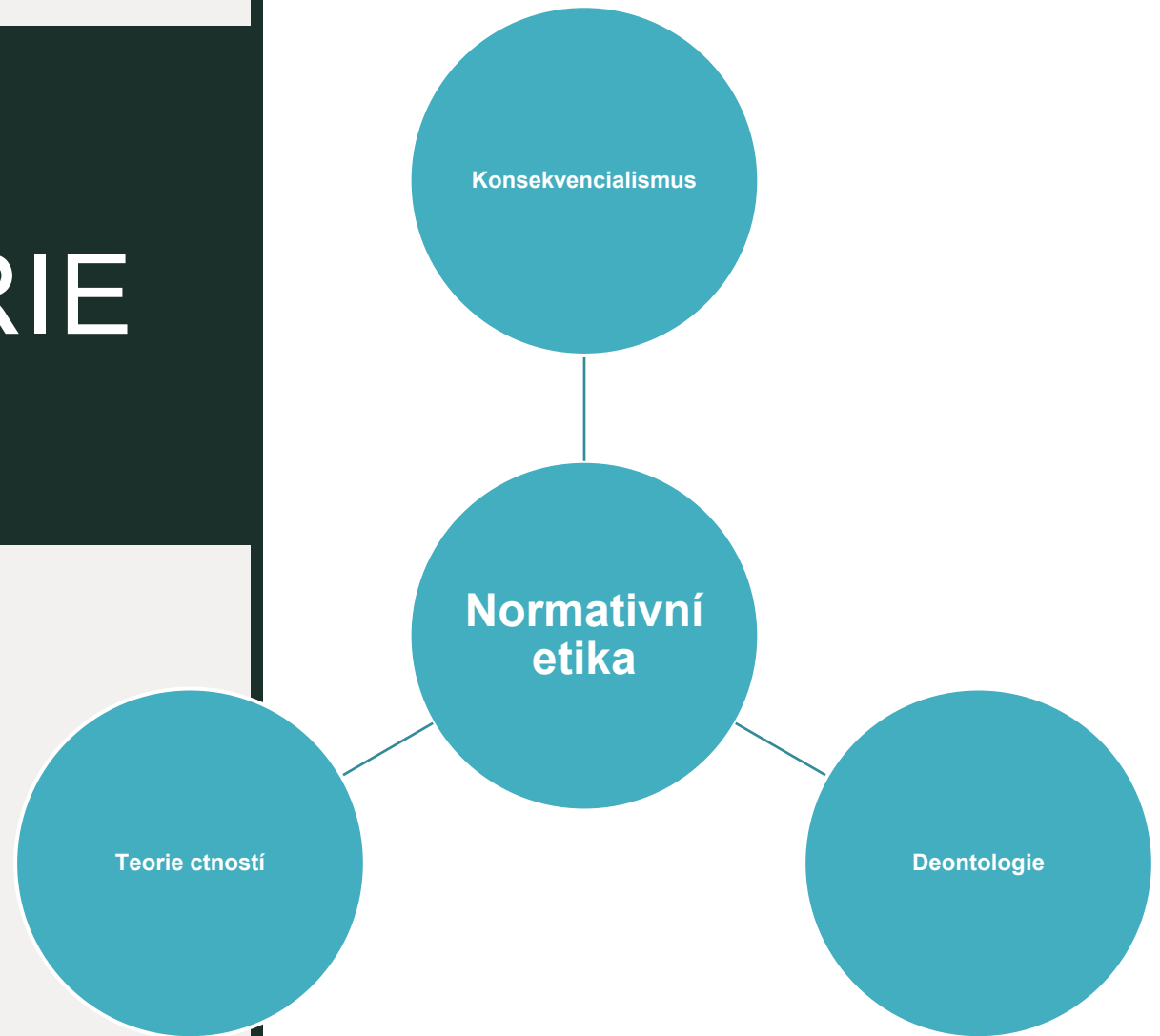
Negativní:
nesplněné
touhy



PLNOHODNOTNÝ LIDSKÝ ROZVOJ



ETICKÉ TEORIE



KONSEKVENCIÁLISMUS

Jádro teorie

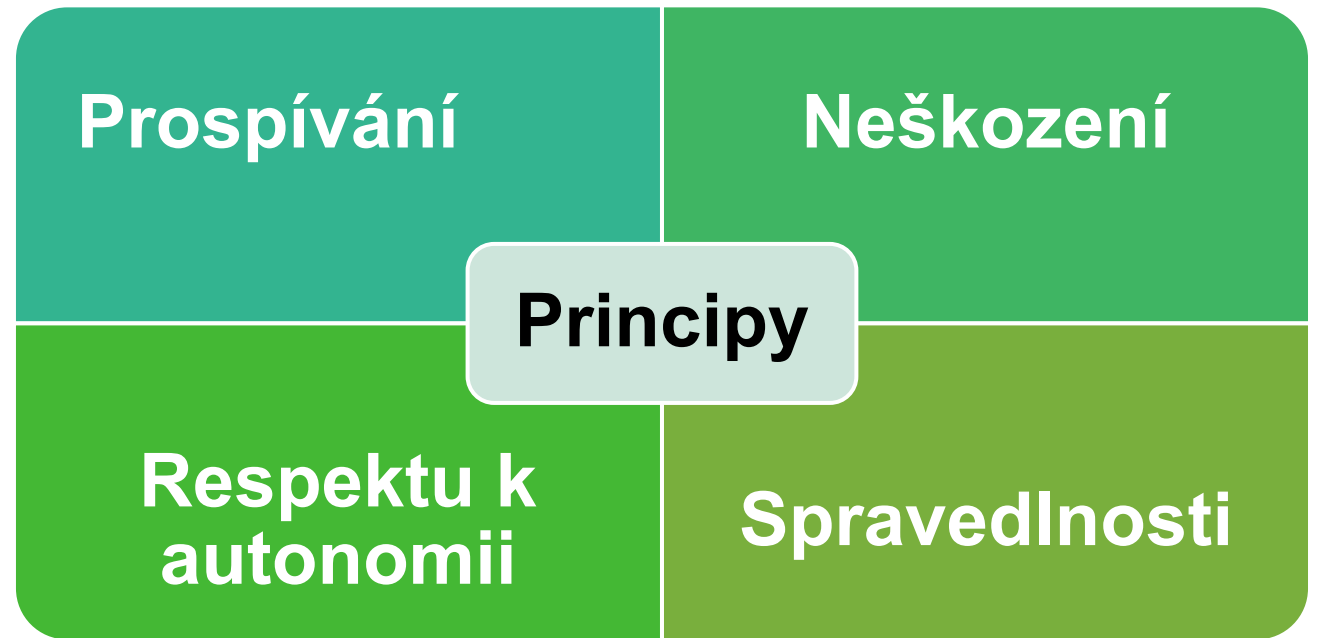
Důsledky

Interpretační
klíč

UTILITARISMUS



Principlismus



Etika ctností

Moudrost a poznání
(rozumové zdatnosti nezbytné pro korektní aplikaci zdatností morálních).

- Kreativita.
- Zvědavost.
- Otevřenost jiným názorům.
- Láska ke vzdělání.
- Moudrost (poznání sebe sama a svých limitů).



Odvaha (schopnost naplňování cílů tváří v tvář obtížnostem).

- Statečnost.
- Vytrvalost.



Lidskost (mezilidské vztahy).

- Altruismus (ochota pomáhat jiným).



Spravedlnost (ctnost regulující vztahy ve společnosti).

- Férovost.



Umírněnost
(schopnost čelit pokušení k extrémním volbám a jednání).

- Sebekontrola.

See you soon!

