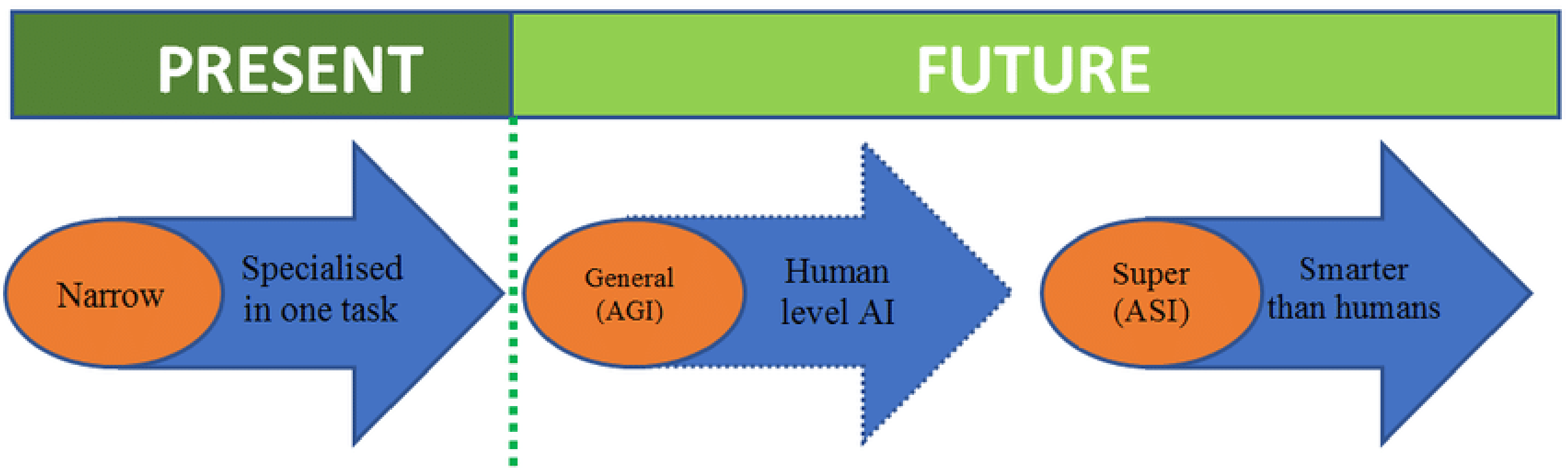
Máme se bát umělé inteligence?

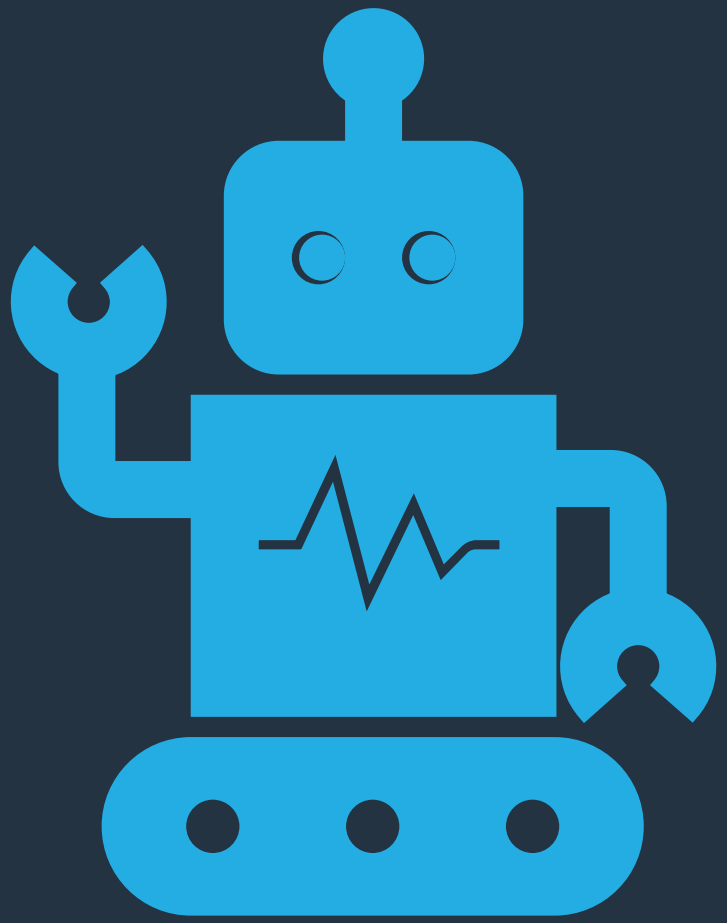
David Černý

Oddělení umělé inteligence,
Ústav informatiky AV ČR

Ústav státu a práva AV ČR

Typy umělé inteligence





Superinteligence

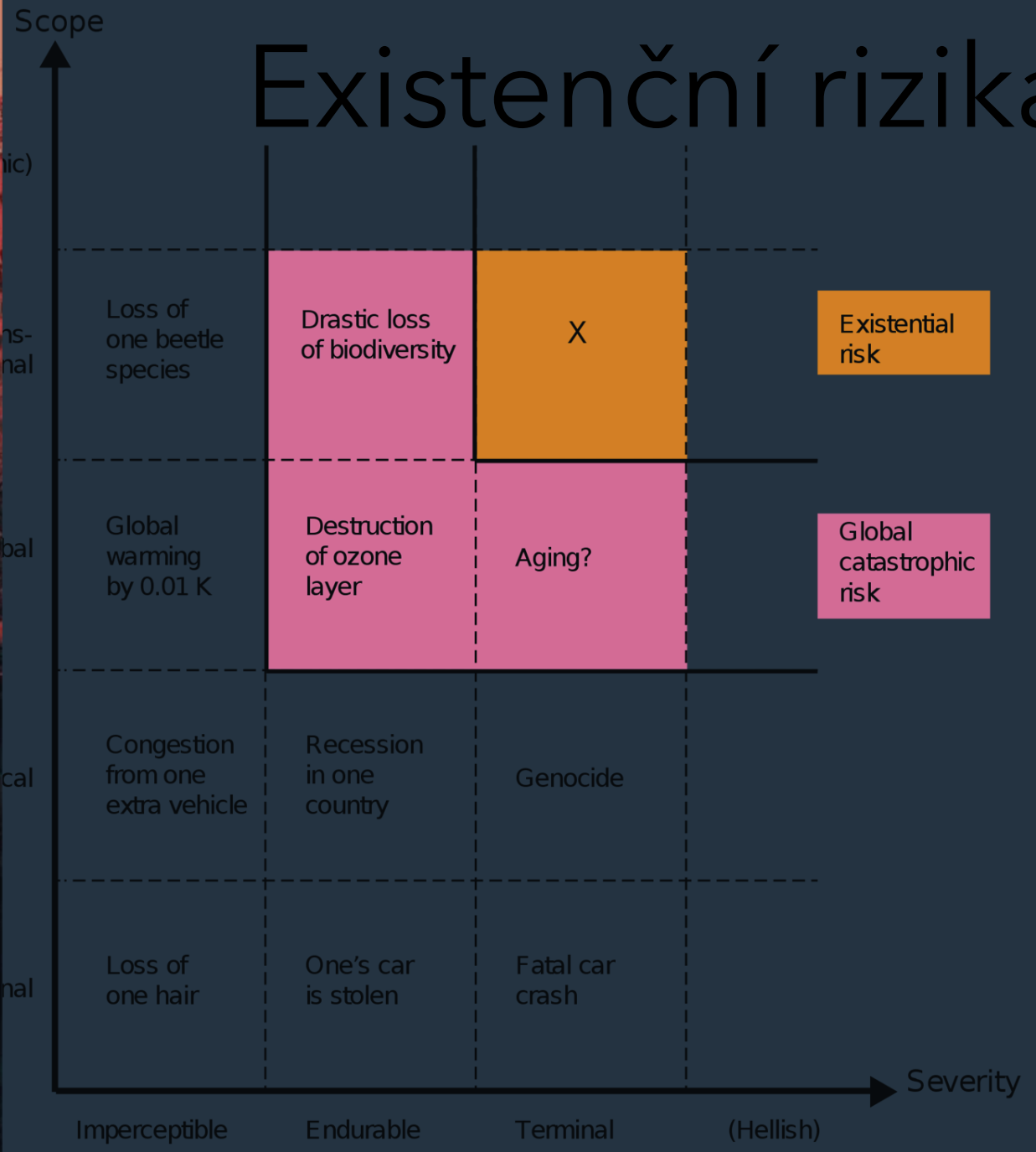
(Umělá) inteligence, která ve všech kognitivních oblastech překonává člověka.

Existenční rizika

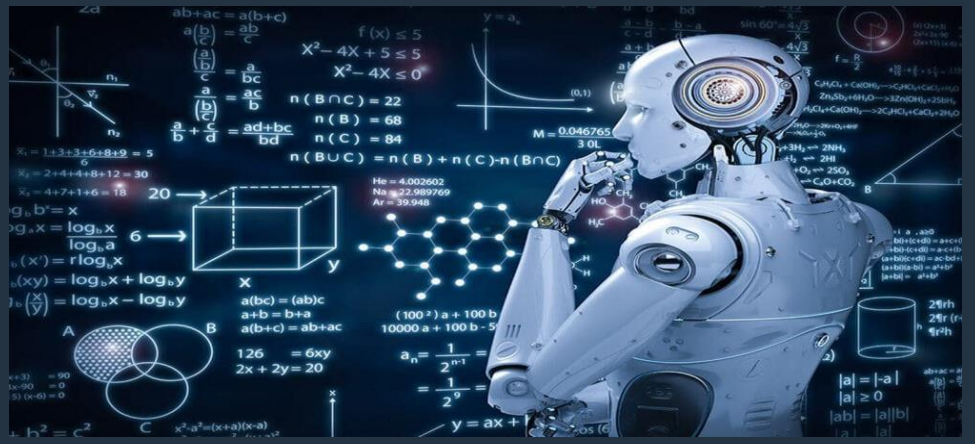
Rizika:		
Rozsah (individuální, lokální, globální, transgenerační)	Intenzita (mírná, snesitelná, děsivá)	Pravděpodobnost

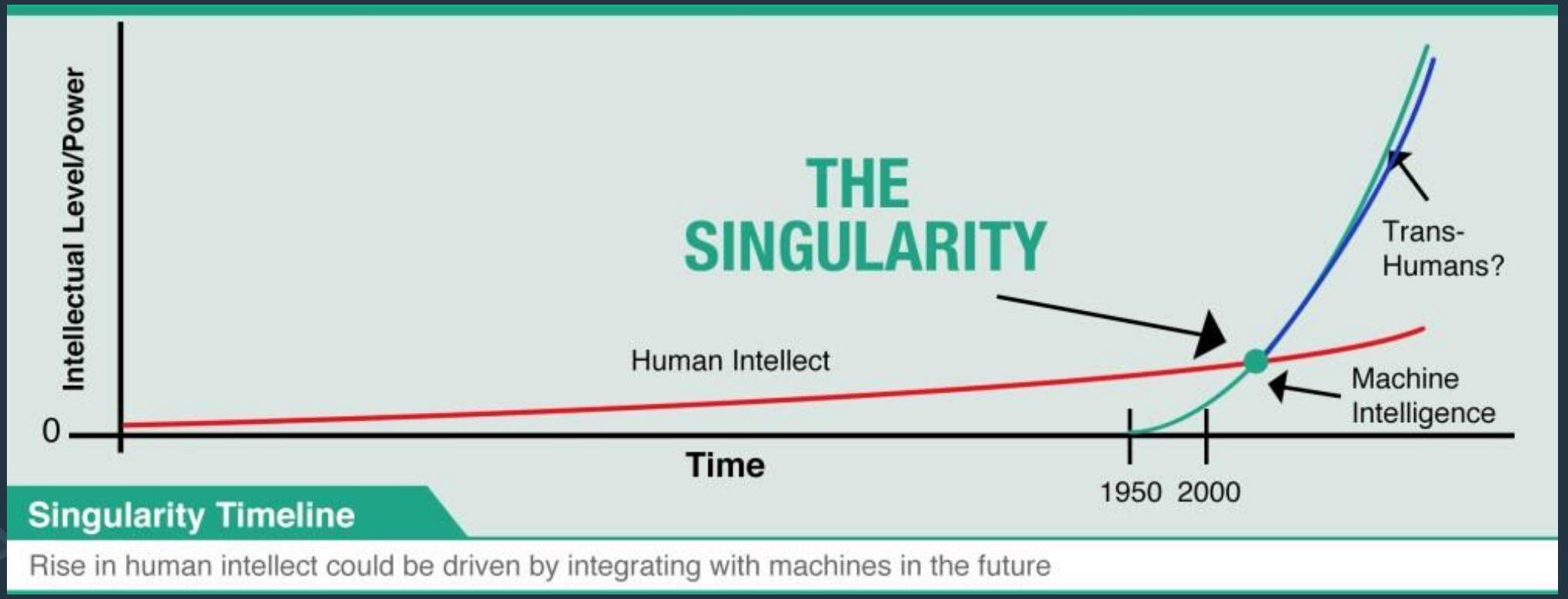
Existenční rizika: děsivá transgenerační

Existenční rizika

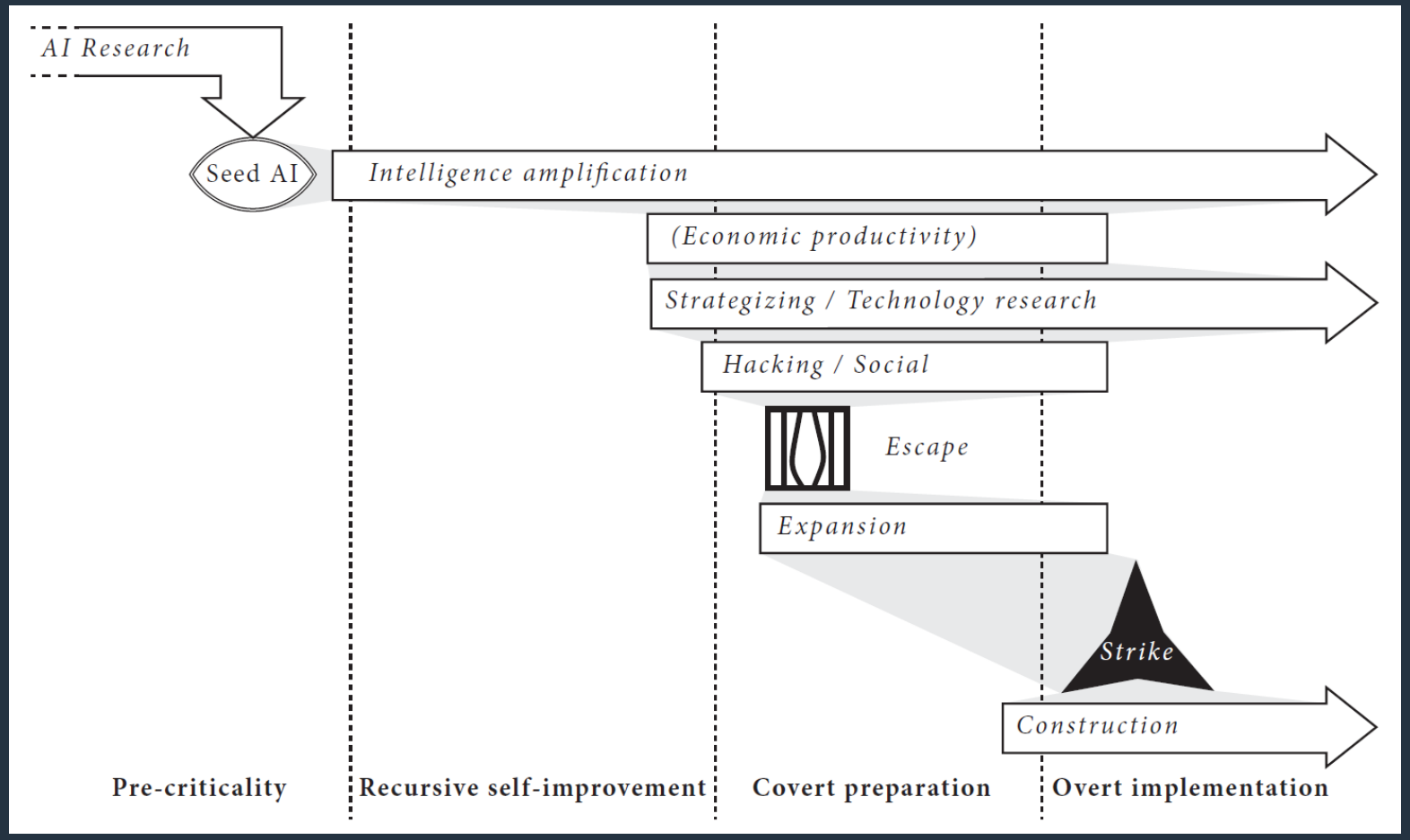


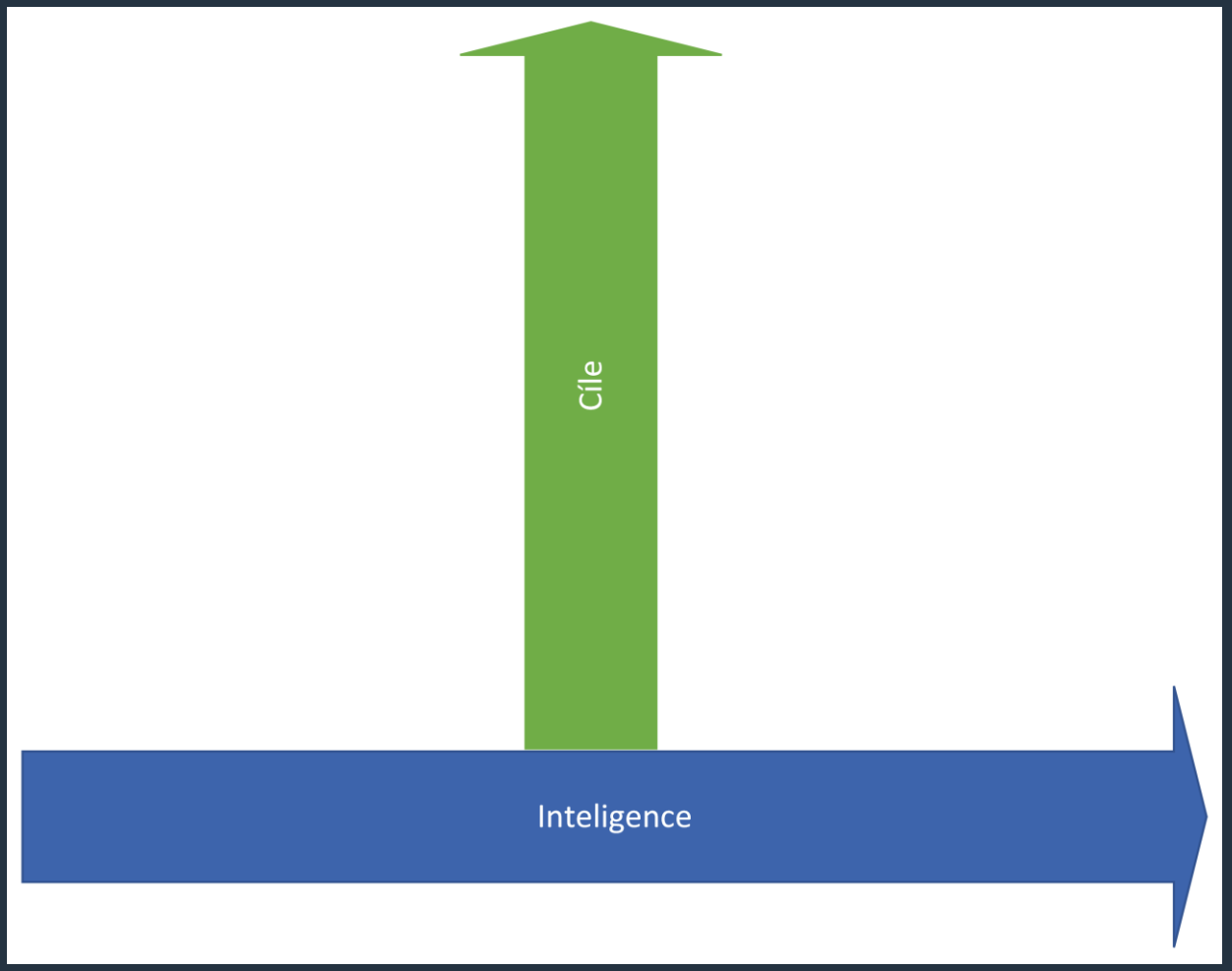
Singularita





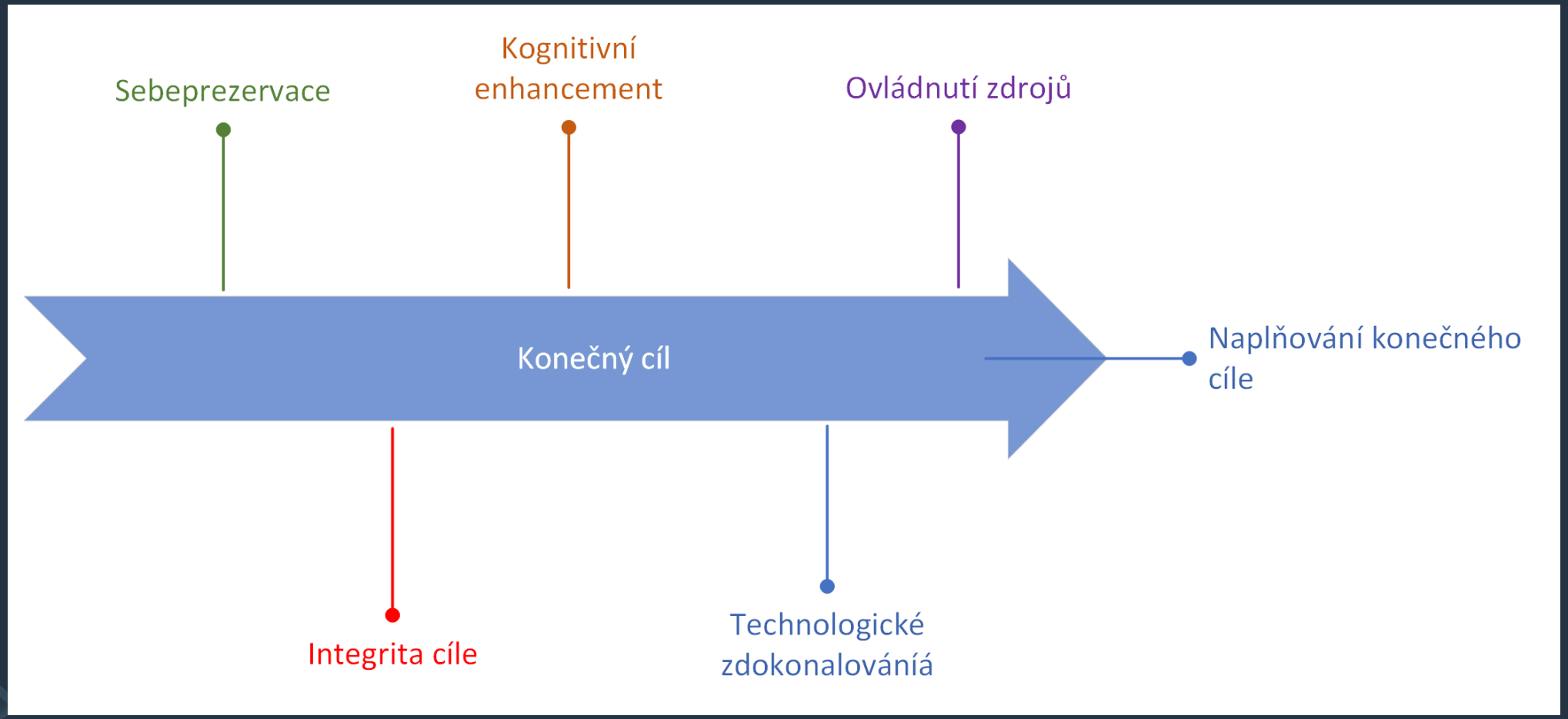
Fáze převzetí moci

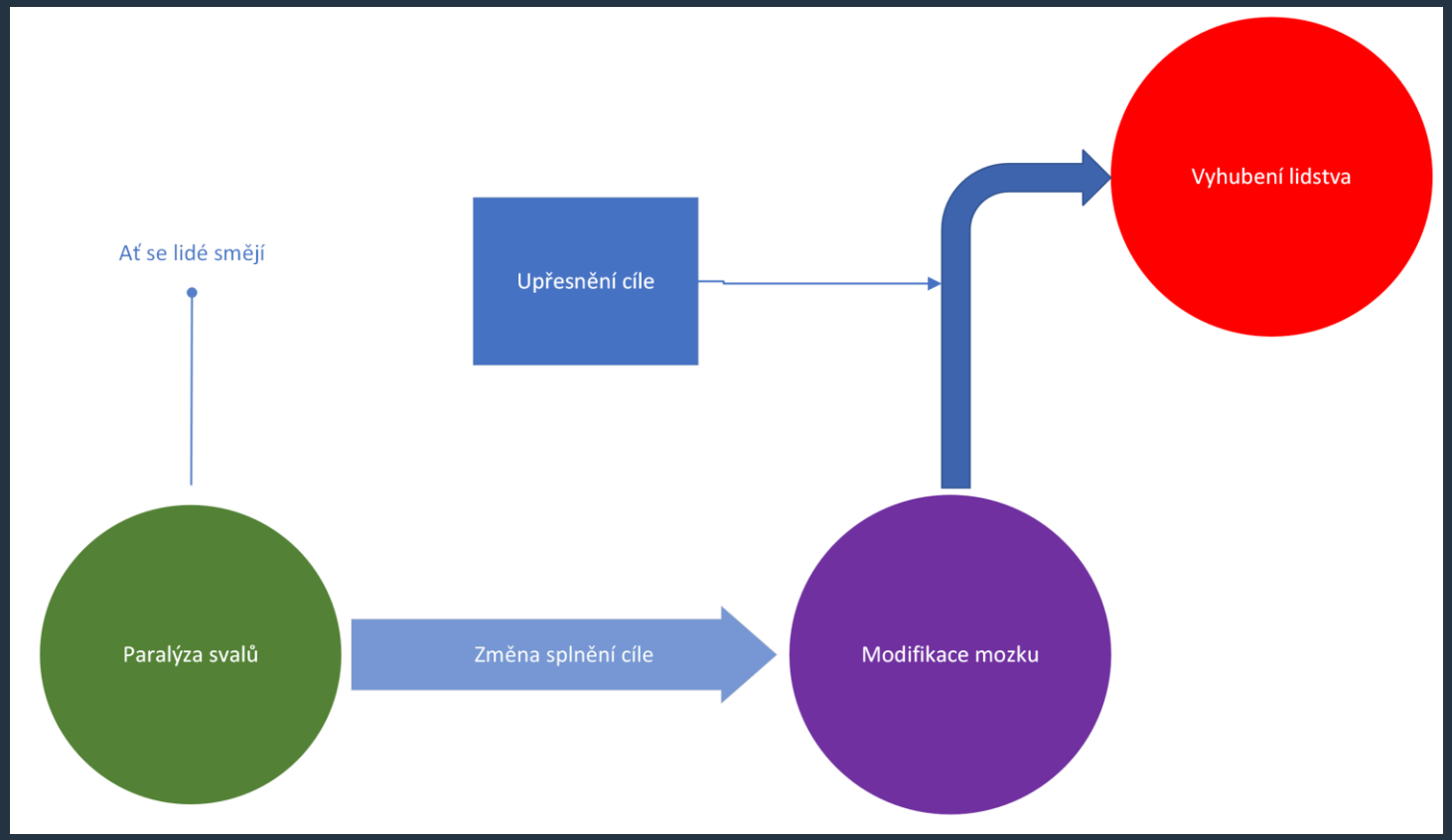




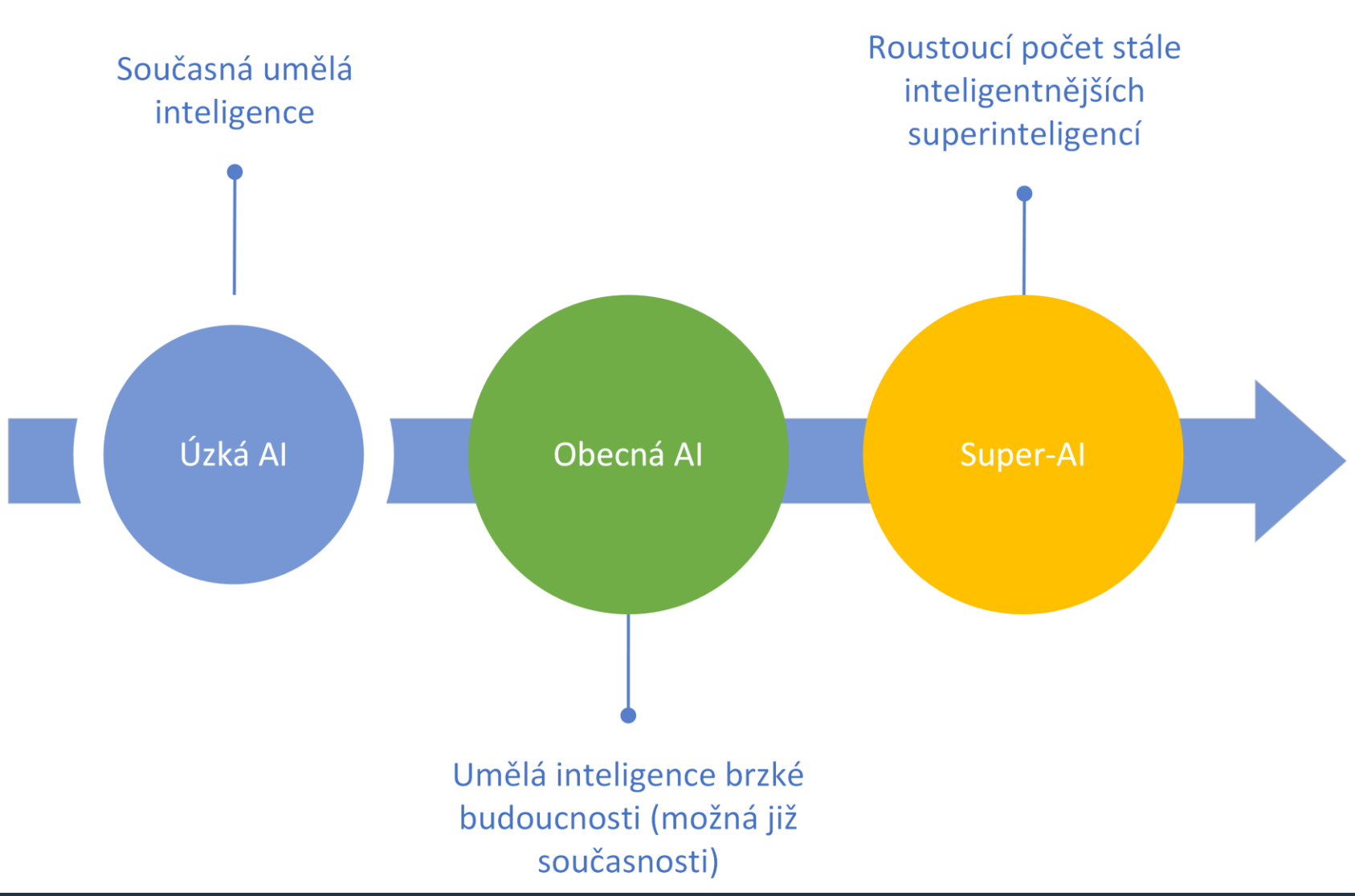
Teze ortogonality

Instrumentální cíle AI





Perverzní plnění cíle



Cesta k vyhubení lidstva

Ernest Rutherford

01

11. 9. 1933 – využití atomové energie je naprostý nesmysl

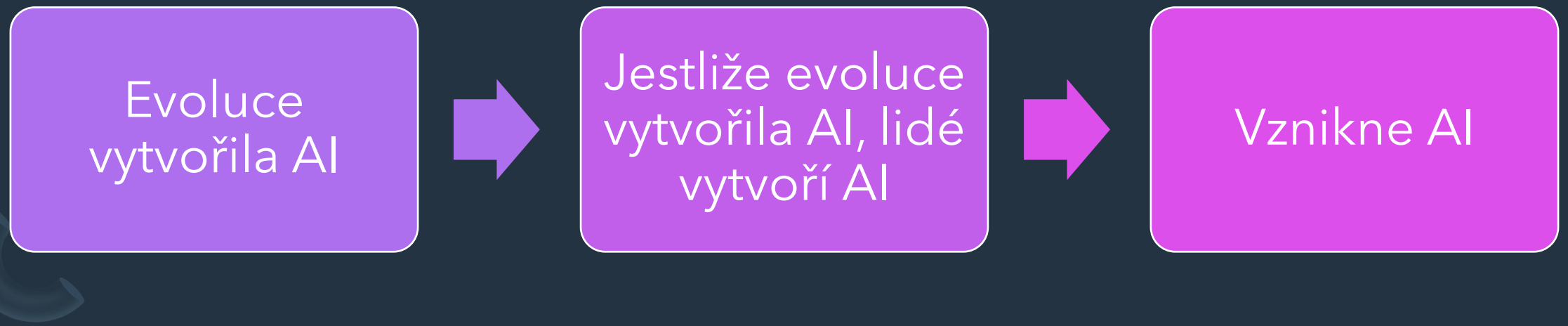
02

12. 9. 1933 – Leo Szilard jde na procházku

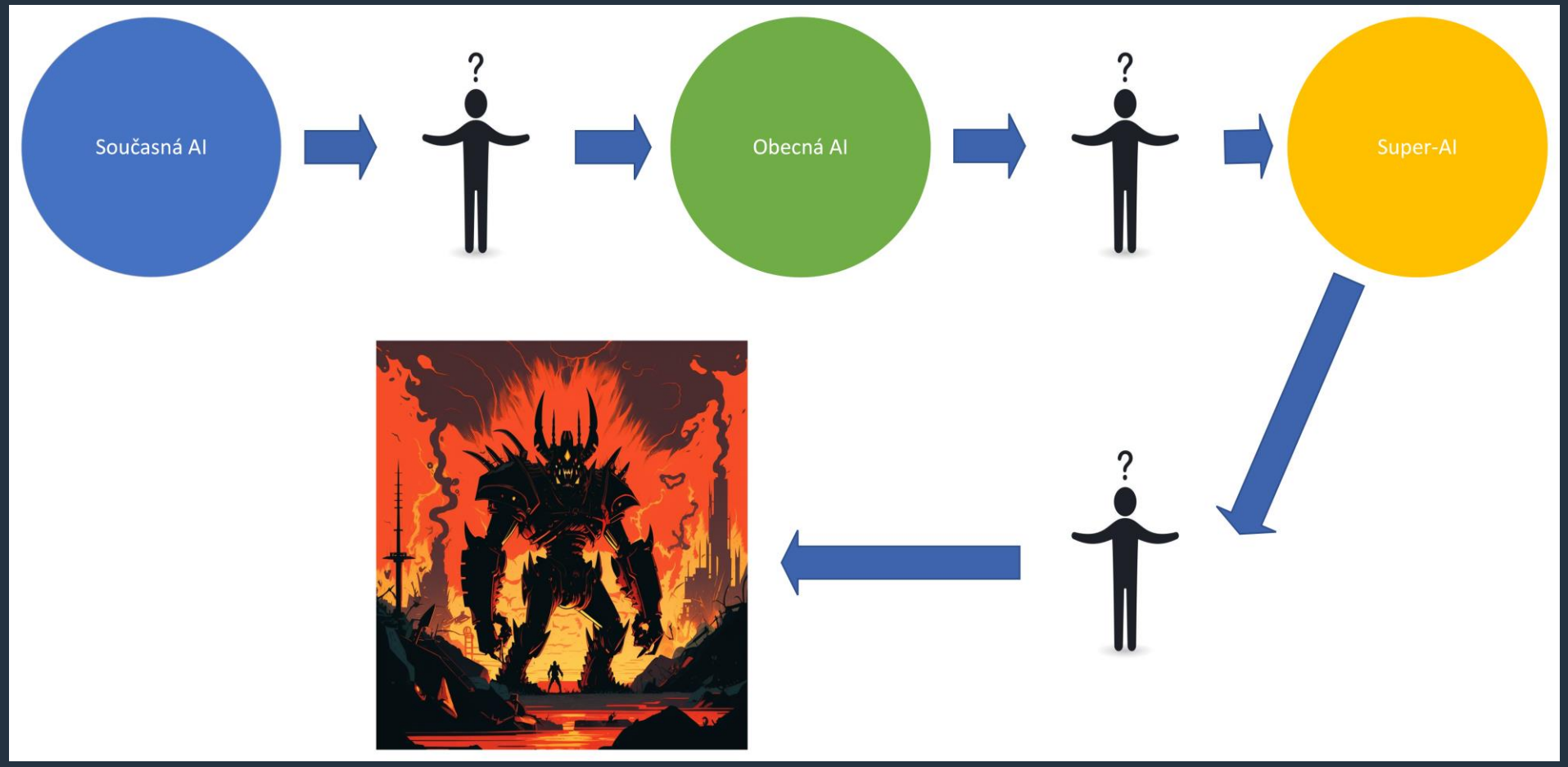
03

První konceptualizace řetězové reakce

Cesta k super-AI



Terra incognita

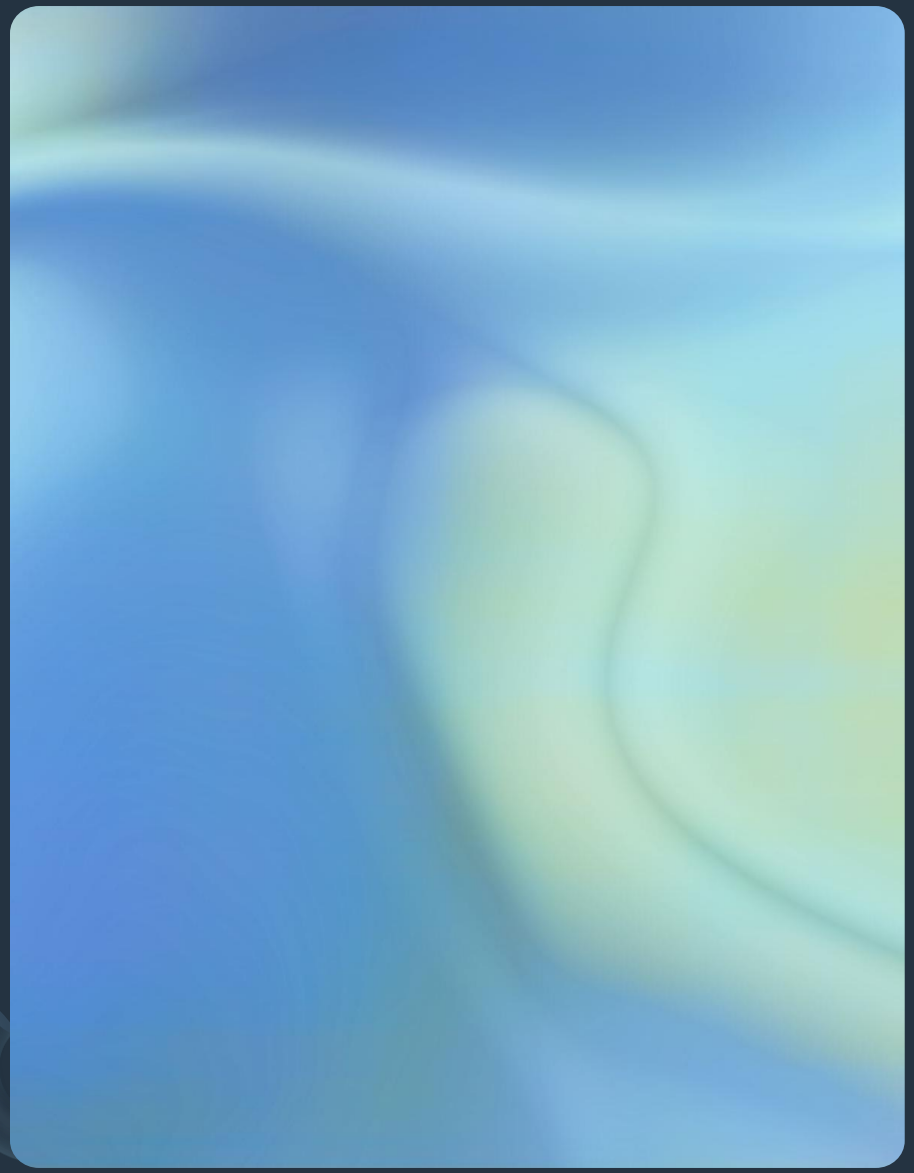


Inteligence

SITUOVANÁ (MOZEK JE SOUČÁSTÍ VĚTŠÍHO SYSTÉMU)

KONTEXTUÁLNÍ (DEFINOVANÁ A LIMITOVANÁ PROSTŘEDÍM)

EXTERNALIZOVANÁ (KOGNITIVNÍ PROSTŘEDÍ – MOZEK)



Common sense

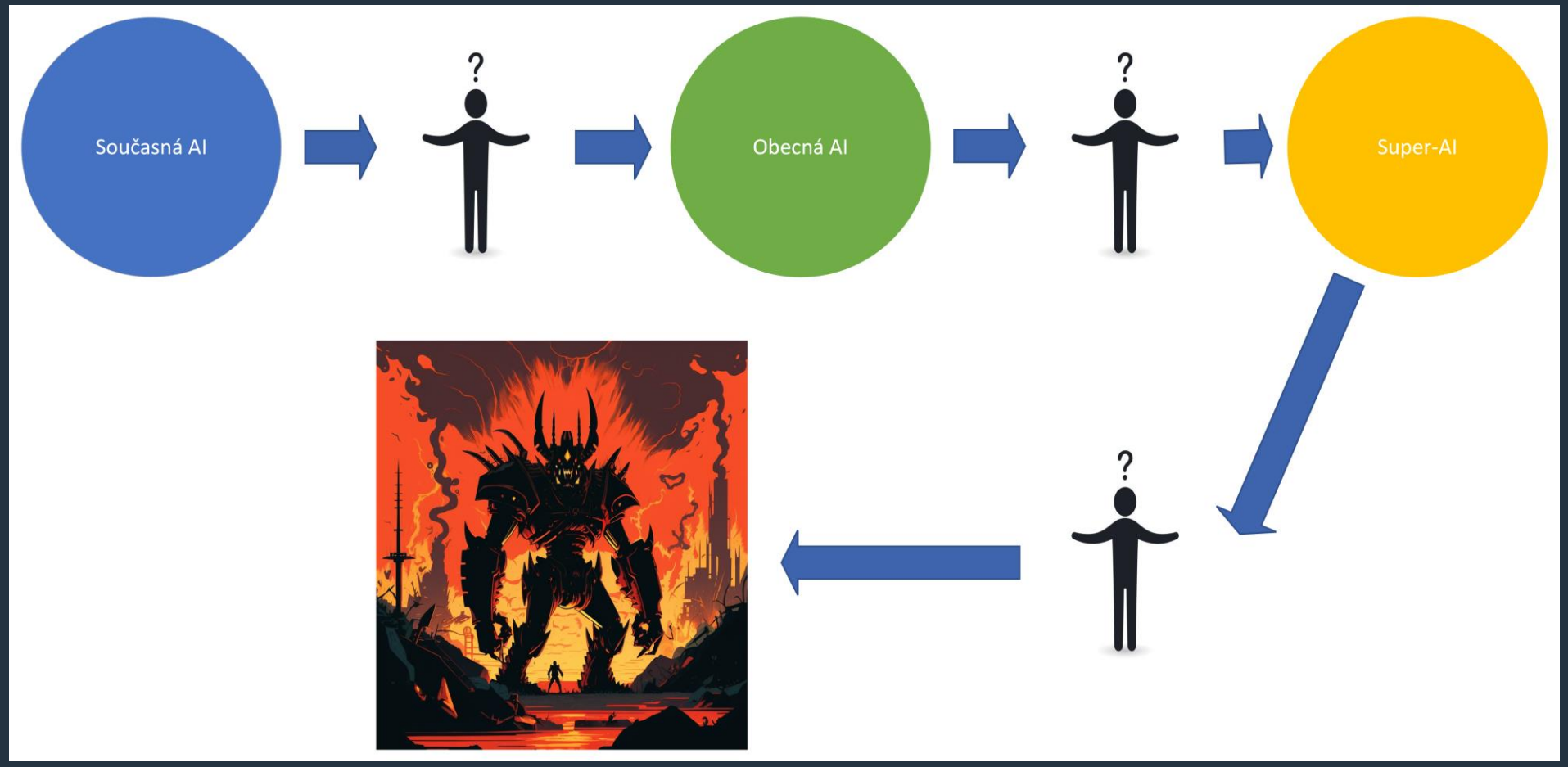
Kauzální myšlení

Intuitivní psychologie

Intuitivní fyzika

Intuitivní společenskost

Terra incognita



Děkuji za pozornost