

# Typy filozofických otázek vztahující se k umělé inteligenci

## Ontologické otázky

(ontologie = učení o bytí, o jeho nejobecnějších určeních a pojmech [<http://slovník-cizich-slov.abz.cz>])

Týkají se povahy bytí předmětu zkoumání, tj. v souvislosti s umělou inteligencí zejména:

- Jaký je rozdíl mezi přirozeným a umělým?
- Co je to mysl, vědomí, inteligence, subjekt?
- Jaký je vztah mysli a těla (mind-body problem)?

## Epistemologické otázky

(epistemologický = týkající se vědeckého poznání a poznávání, též noetický, gnozeologický [<http://slovník-cizich-slov.abz.cz>])

Týkají se našeho poznání, tj. co vůbec můžeme vědět o přirozeném myšlení, jakými cestami se o něm dozvídat a jaké poučení lze čekat od umělého modelování.

## Metodologické otázky

(metodologie = nauka o metodách, zejm. vědeckého bádání [<http://slovník-cizich-slov.abz.cz>])

Jaké jsou motivační zdroje a paradigmata umělé inteligence?

## Futurologické otázky

(futurolgie = teorie budoucnosti, výzkum možných budoucností [<http://slovník-cizich-slov.abz.cz>])

Kam se ubírá vývoj umělé inteligence?

Jak umělá inteligence změní svět?

## Etické otázky

(etika = teorie mravnosti; soustava názorů na mravnost, systém mravních norem; morálka, mravnost [<http://slovník-cizich-slov.abz.cz>])

Je to všechno dobré?

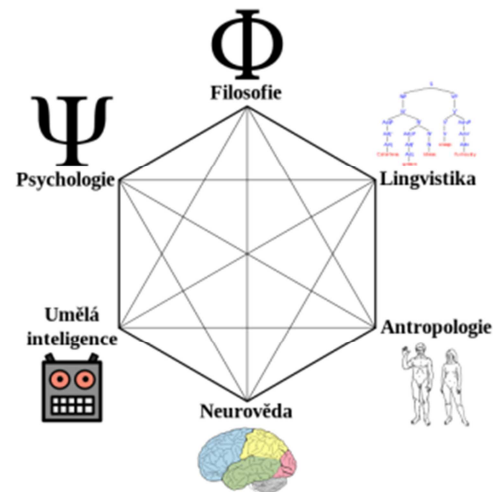
## Motivační zdroje umělé inteligence

- **motivace aplikační, inženýrská** (v současné době převládající) – to, co bylo v rámci předmětu ZUI probíráno dosud:
  - Umělá inteligence je **věda** o vytváření strojů nebo systémů, které budou při řešení určitého úkolu užívat takového postupu, který – kdyby ho dělal člověk – bychom považovali za projev jeho inteligence (Minsky).
  - Umělá inteligence je **vlastnost** uměle vytvořeného systému, který má schopnost rozpoznávat předměty a jevy, analyzovat vztahy mezi nimi a tak si vytvářet modely světa, dělat účelná rozhodnutí a předvídat jejich důsledky, řešit problémy včetně objevování nových zákonitostí a zdokonalování své činnosti (Kotek).
- **motivace badatelská** – snaha dozvědět se co nejvíce o povaze mentálních procesů, především těch, které známe u člověka (o to nám půjde v této části přednášek)
- **motivace matematická** – čistě teoretické studium matematických a logických systémů odvozených z metod umělé inteligence
- **přirozená lidská hravost** – co všechno dovedeme sestrojít a naprogramovat

## Umělá inteligence a kognitivní vědy

Kognitivní vědy lze chápat jako sdružení umělé inteligence (tak, jak je chápána z hlediska její badatelské motivace), psychologie, neurověd, lingvistiky, filozofie mysli a antropologie pod jedním pojmem.

**Kognitivní vědy** studují všechny formy *lidské* inteligence, od vnímání a jednání (konání) až po řeč a myšlení.



## Co je předmětem zájmu badatelské umělé inteligence (obecněji kognitivních věd)?

- Předmětem zkoumání je „přirozené myšlení“, mentální procesy. Co to ale je mentální proces? Co je to mysl? Co je duše, psychika, inteligence, intelekt, vědomí?
- V běžné mluvě tato slova používáme a rozumíme jim, ale pokud je začneme brát jako odborné termíny, začnou vědci (zejména z oblasti typicky deduktivních věd, jako je matematika) požadovat definici, co se jimi vlastně míní. V humanitních oblastech není definice pojmu nutná, předpokládá se, že jednotlivé pojmy se mění až s naším poznáním.

## Obecné přístupy ke zkoumání přirozené mysli v kognitivních vědách

- **Cesta vnitřního prožívání** (inner experience, introspekce a fenomenologie) – je to „vnitřní“ pohled sám na sebe.
  - Co ale znamená „vnitřní“?  
Rozlišení na vnitřní a vnější vychází z metafory nádoby – odkud se na nádobu díváme?  
Když ale zkoumáme sami sebe, jsme uvnitř nádoby nebo vně?
    - **introspekce** – pozoruji sama sebe jakoby zvenku (vystoupím ze sebe)
    - **prožívání** – prožívám sebe sama zevnitř, nepozoruju se přitom
  - U tohoto přístupu je problém s objektivizací.
- **Cesta přírodních věd** – přírodní vědy mají předmět svého studia před sebou, mohou jej pozorovat, ohmatávat, měřit, atd.
  - Mentální stavy ale nejsou hmatatelné → přírodovědec zkoumá buď:
    - vnější projevy mysli (chování jednotlivců, skupin,...) – jde o pohled „shora“, nebo
    - „biologický substrát“ mysli, tj. mozek – jde o pohled „zdola“. Už víme, jak vypadá mozková tkáň, co se v ní odehrává v různých situacích, známe funkční oblasti mozku, ale *jak to vše souvisí s myšlením?!?!* Je myšlenka průvodní jev nějaké fyzické příhody v mozku? Nebo naopak ta fyzická příhoda je průvodním jevem toho, že se zrodila myšlenka?
  - V souvislosti s mozkiem existuje pořád celá řada záhad:
    - Jak je možné, že každý jednotlivý mozek vůbec funguje a že dovede věci, které se nám zdají tak překvapivě inteligentní?
    - Jak je možné, že mozek jednotlivce dospěje k této dokonalosti během pouhých několika let a to takřka z ničeho (z jediné buňky)?
    - Jak je možné, že evoluce vůbec něco takového vynalezla?
    - Proč jsou pro nás tyto otázky stále záhadami, když už toho o mozku víme tolik?
  - Poznámka: Rozlišení „shora“ a „zdola“ je pouze metaforické
- **Cesta umělého modelování** – experimentuje se s uměle vytvořenými modely (matematickými, počítačovými, fyzikálními, fyzickými). Hlavním reprezentantem této cesty je právě umělá inteligence.
  - Uměle vytvořené modely se často simulují na počítači, ale zároveň lze počítač chápat jako model lidské mysli (počítačový funkcionalismus, silná umělá inteligence)
  - Existují 2 typy modelů:
    - **věrný model** – fyzický nebo teoretický (zpravidla matematický) konstrukt, snažíme se jím co nejlépe napodobit vybrané vlastnosti modelovaného objektu.

- **metaforický model** – připodobňuje něco neznámého tomu, co již známe, nebo čemu již rozumíme, přičemž i odlišnost od modelované skutečnosti nám pomáhá něco objasnit či vysvětlit. Za metaforický model lze považovat i počítač, kdy se hledají paralely mezi činností počítače a myšlením, počínaje logickými prvky a konče sekvenčním uspořádáním výsledného procesu.
- I u modelování lze hovořit o přístupu „shora“ a „zdola“:
  - U přístupu „shora“ jsou východiskem symbolické, na reprezentaci založené procesy myšlení. Rerezentantem tohoto přístupu je tzv. **tradiční umělá inteligence**.
  - Pro přístup „zdola“ je třeba stanovit **základní úroveň analogie**, „pod“ níž již není rozhodující, zda si prvky modelu zachovávají či nezachovávají strukturní, materiální nebo formální podobnost s odpovídajícími prvky modelované skutečnosti. Stačí, že mají shodnou tu část chování, která je směrodatná v rámci znalosti modelovaného systému.

## Paradigmata v umělé inteligenci

- **tradiční umělá inteligence** (používá se též pojem logicko-symbolické, symbolicko-reprezentační, algoritmické nebo komputacionalistické paradigma; užívá se též označení GOF AI = Good Old-Fashioned Artificial Intelligence)
- **konekcionismus** (neuronové sítě) a **evoluční princip** (genetické algoritmy)
- **distribuovaná umělá inteligence** (multiagentní systémy)

## Tradiční umělá inteligence

(patří sem vše, o čem jsme se bavili na předchozích přednáškách)

Jaký je rozdíl mezi umělým a přirozeným? Jestliže se shodneme na intuitivním chápání pojmu „přirozený“, pak o nějaké věci (činnosti, vlastnosti) můžeme říct, že je „umělá“, když:

1. existuje nějaká **přirozená** věc (činnost, vlastnost), logicky připouštějící duplikaci (květina, jezírko, drahokam, ledvina, myšlení, vnímání, rozhodování apod.)
2. existuje **záměr** vytvořit duplikát té přirozené věci
3. došlo k **provedení záměru**, tj. proběhl intencionální proces vedoucí od záměru k jeho realizaci (v podstatě tedy musíme být schopni definovat konkrétní cíl, co chceme vytvořit)

U **mentálních procesů** lze rozlišit 2 komponenty: **performační** (to je to, co je popsitelné zvnějšku, tedy to, co je na procesu objektivní) a **fenomenální** (to je to, co subjekt „vnitřně“ prožívá).

Tradiční umělá inteligence se snaží modelovat performační složku.

## Silná umělá inteligence (počítačový funkcionalismus)

Je to názor, který lze vyjádřit tezí „Povaha myslí je algoritmická, přičemž není podstatné, v jakém médiu jsou algoritmy (programy) implementovány“. Tj. nezáleží na tom, zda tímto médiem je mozek, počítač, či armáda Číňanů. Počítače by proto mohly (dokonce musely) mít myšlenky, pocity, porozumění apod. Stačí, aby do nich byl implementovaný vhodný počítačový program s vhodnými vstupy a výstupy.

## Slabá umělá inteligence

Snaží se pouze **modelovat** mysl, popř. její dílčí projevy (např. logické uvažování) na logicko-symbolické úrovni.

## Jak ověřit, zda stroje mohou myslet

### Turingův test

Představme si 2 komory, v jedné je stroj, v druhé člověk. Vnější experimentátor má na základě dialogu s oběma komorami rozhodnout, kdo je kde. Stroj se bude přitom vydávat za člověka, člověk též. Otázka zní, zda experimentátor bude (v průměru) rozhodovat stejně úspěšně, jako kdyby v jedné místnosti byl muž a v druhé žena vydávající se za muže.

Test nebere v potaz fenomenální složku myšlení. V podstatě se testuje pouze to, jak dobře umí stroj předstírat, že není stroj.

## Argument proti počítačovému funkcionalismu

### Čínský pokoj

V uzavřené místnosti sedí člověk, který nerozumí čínsky, a pomocí manuálu (psaném v jazyce, kterému rozumí) vybírá jedny čínské znaky jako odpovědi na jiné čínské znaky, které mu do místnosti vsouvá rodilý Číňan. Manuál je přitom přesným přepisem toho, jak vést v čínštině smysluplný dialog. Rodilý Číňan musí zákonitě dojít k závěru, že člověk sedící v místnosti rozumí čínštině, což ale není pravda. Když tedy v místnosti místo člověka bude počítač, který bude podle stejného manuálu „smysluplně“ reagovat na zadané vstupy, neznamená to, že počítač rozumí čínštině.

## Emergentismus

U tradiční umělé inteligence jsme předpokládali, že musí existovat konkrétní cíl, co chceme vytvořit. Jestliže budeme uvažovat modelování „zdola“, můžeme se na nějaké „vyšší“ úrovni dostatečně složitého systému setkat s vlastnostmi, procesy či jevy, se kterými jsme předem nepočítali a které předem nedovedeme popsat. Takové vlastnosti, procesy a jevy se nazývají **emergentní**. Otázka je, zda fenomenální složka myšlení může být takovouto emergentní vlastností umělého média.

V umělé inteligenci se uplatňují 2 strategie, ve kterých se projevuje emergence: konekcionismus (neuronové sítě) a evoluční přístup (genetické algoritmy).

### Konekcionismus

Je založen na spojování velkého množství základních relativně jednoduchých funkčních prvků s jednoduše definovaným chováním, ale výsledný celek má díky emergentismu kvalitativně naprosto odlišné chování (analogie s mozkem, který je tvořen sítí vzájemně propojených „jednoduchých“ neuronů).

### Evoluční přístup

Je inspirován přírodovědeckou evoluční teorií a na základě selekčního tlaku se snaží prosazovat z daného hlediska „nejschopnější jedince“ (jedincem může být robot, algoritmus, ...). Uplatňují se zde principy náhodné křížení a mutace.

## Zjednávací princip

Jak v případě tradiční umělé inteligence, tak v případě emergentismu se předpokládá, že umělý systém vnímá (a tedy „má“) svět shodný s naším. V případě zjednávacího přístupu je stroj plně autonomní systém, který si „žije svým vlastním životem“. Příkladem jsou tzv. **reaktivní agenti** v robotice. Určité množství agentů může vytvářet vyšší jednotky (jakási společenství, kolonie apod. – mluvíme o tzv. multiagentních systémech). Předpokládá se, že takto vytvořený systém by měl mít vyšší inteligenci než jednotliví členové společenstva (podobně jako mraveniště nebo včelstva).

## Futurologické otázky UI

V současnosti se futurologové zabývají hlavně dvěma otázkami:

1. Jak se nám bude žít mezi dokonale inteligentními roboty?
2. Jak bude vypadat svět vybavený globálním superinteligentním kyberprostorem (*cyberspace*), jehož zárodkem je dnešní internet (Průmysl 4.0)?

Každá futurologická myšlenka je založena na určitých předpokladech:

- Explicitní předpoklady – futurolog z nich zpravidla vychází a zakládá na nich své prognózy; jde zpravidla o extrapolaci ze současného stavu a současných znalostí
- Implicitní předpoklady – futurolog si je často ani sám neuvědomuje

### Příklad:

Kevin Warwick (anglický kybernetik) : *Zatímco dnes jsme my, lidé, díky naší inteligenci, dominantní formou života na Zemi, je možné, že již relativně brzo budou stroje inteligentnější než lidé. Pak se stroje stanou dominantní formou života na Zemi.*

Přitom podle Warwicka je pro inteligenci rozhodující počet a hustota neuronů, kvalita vzájemných vazeb mezi nimi a s tím související schopnost se učit. V takto definované „inteligenci“ nás stroje opravdu brzo předhoní, zatímco my nemůžeme svoje neurony zmnožit a přimět je, aby pracovaly lépe než teď.

Warwickovy implicitní předpoklady:

- 1) *dnes jsme my, lidé, dominantní formou života na Zemi*
- 2) *naše dominance je dána naší inteligencí*
- 3) *Pro inteligenci je rozhodující počet a hustota neuronů, kvalita vzájemných vazeb mezi nimi a s tím související schopnost se učit*

## Etické otázky UI

Opravdu chceme vytvořit uměle inteligentní stroj, jehož podoba, chování, vlastnosti apod. budou nerozeznatelné od člověka?

