Umělá inteligence a filosofie

referát do předmětu PB016 – Úvod do umělé inteligence, podzim 2005

Pokud je proces lidského myšlení možné realizovat výpočtem, může k tomu dojít i dostatečně dlouhým přemísťováním kuliček na kuličkovém počítadle?

Vztah mezi umělou inteligencí a filosofií

kognitivní vědy

– *zabývají se studiem myšlení, inteligence, vědomí...*
– *pojmy lze obtížně definovat, popis je složitý*
– *různé metody (hypnosa, introspekce, biologický popis...)*

umělá inteligence

– *umělé modelování lidských mentálních procesů*
=> *jedna z metod kognitivních věd*
=> *objekt zkoumání kognitivních věd*

Kde jsou hranice umělého modelování mysli?

Pohled do historie

1750 – Julien Offray de La Mettrie

– *„Můžeme směle prohlásit, že člověk je pouhý stroj.“*

1. pol. 19. století – Ada Lovelace

– *odmítá strojové myšlení*

– *myšlení je spojeno s původností x počítač pouze modifikuje vstup*

námitky:

– *co je to původnost, je člověk „původní“?*

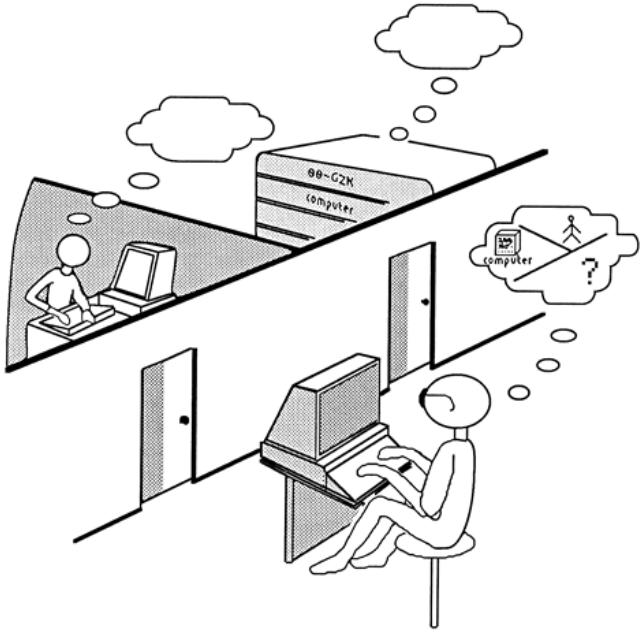
– *např. báseň = permutace slov (= výsledek určité funkce)*

– *na lidské chování lze pohlížet na modifikaci vstupů a reakce na ně*

Turingův test

– 1950 Alan M. Turing – text Computing Machinery And Intelligence
– kritérium, které musí stroj splnit, aby mohl být označen za inteligentní

– imitační hra – 2 uzavřené místnosti, soupeří spolu člověk a stroj
– pokud rozhodčí na základě dialogu není schopen zjistit, ve které místnosti je stroj, pak stroj označíme za inteligentní



– mnoho filosofů mysli test považuje za příliš „měkký“
– přesto se nedaří jej splnit

– od r. 1991 každoročně soutěž o splnění Turingova testu – Loebnerova cena

http://www.loebner.net/Prizef/loebnerprize.html

Loebnerova cena


TUESDAY, NOVEMBER 5, 1991 Page C10

C.M. Haridi for The New York Times

Hugh Loebner, New York philanthropist, is offering $100,000 in contest supported by National Science Foundation and Sloan Foundation to determine whether computer programs can imitate human responses.

– *cíl = splnění Turingova testu (hlavní cena 100 000 $)*
– *každoroční cena pro nejlépe hodnocený program – 2000 $*

pravidla:
– *ze všech účastníků (programů) jsou vybráni 4 finalisté*
– *ve finále se realizuje Turingův test:*
  – *4 rozhodčí, 4 „lidští partneři“ programů*
  – *4 kola po třiceti minutách – každý rozhodčí hovoří s každým programem a každým partnerem – bodování podle přesvědčivosti*
  – *sestavena „škála lidskosti“ na základě součtu bodů od všech rozhodčích*
  – *nejlepší program se vždy umístil na 5. místě*

– *vítěz za rok 2005 – program Jabberwacky (R. Carpenter)*
  www.jabberwacky.com – online implementace

Loebnerova cena

– *program Jabberwacky jeví známky osobnosti (smysl pro humor, učení apod.)*

– *neschopnost odpovědět na triviální otázky o okolním světě („Which is larger, a TV screen or Belgium?“)*

U: So, tell me something about you.
J: I am 16 years old.

U: When were you born?
J: I was born in May of 1983.

U: How old are you?
J: 19 years old.

U: So you are a lier.
J: Yes.

Termíny filosofie mysli

Počítačový funkcionalismus (silná umělá inteligence, mechanistická teze)

– *Povaha mysli je algoritmická, přičemž není podstatné, v jakém mediu jsou tyto algoritmy implementovány (mozek, počítač, počítadlo)*

dvě komponenty mentálních procesů

– *performační* – *projevuje se navenek (chováním), můžeme ji popisovat*
– *např. činnost hraní šachů (tahy, váhání)*

– *fenomenální* – *je skrytá, nepopsatelná, víme o ní pouze u sebe*
– *subjektivní prožitek při hraní šachů (baví mě to)*

behaviorismus – *redukce zkoumání lidské mysli na performační složku (Turing)*

emergence – *vznik vlastnosti, která není vysvětlitelná na základě nižší úrovně popisu daného objektu*
– *např. tvar sněhové vločky je emergentní vlastností molekul vody*

Gödelova věta o neúplnosti aritmetiky

„V libovolném systému, který pracuje se základními aritmetickými fakty, existuje otázka, na niž nelze v tomto systému odpovědět (systém je neúplný).“

– *argument proti počítačovému funkcionalismu:*

stroje jsou reprezentanty aritmetických systémů

=> každý stroj je neúplný

=> lidské myšlení je nadřazené strojovému

– *protiargumenty:*

– *i lidské myšlení podléhá neúplnosti: „Odpovíš na tuto otázku chybně?“*

– *velká složitost systému => menší předvídatelnost v chování*

– *možnost vzniku emergentních jevů na nejvyšších úrovních popisu stroje*

Kritika Turingova testu

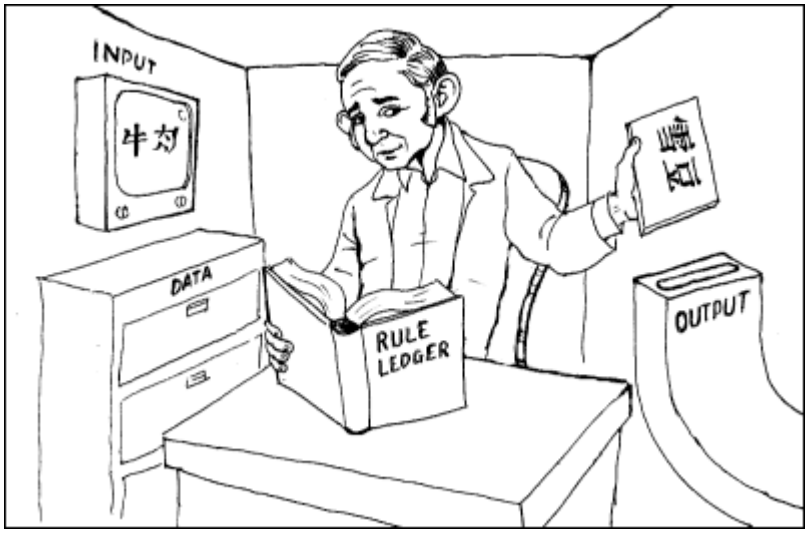
– experiment se omezuje na performační složku myšlení

– malá objektivita – rozhodčím je člověk a výsledek experimentu závisí na jeho důvtipu a nápaditosti

– malá oblast působnosti – stroj umí pouze předstírat, že je člověk
  – člověk neprojevuje inteligenci tím, že předstírá, že je člověk, člověk je člověkem (mohou být lidská a strojová inteligence stejné?)

Chalmersova zombie
– entita, která projde Turingovým testem, ale nemá fenomenální složku osobnosti
  – chceme takovouto bytost označit jako inteligentní?
  – jsou všichni kromě mě Chalmersovými zombiemi?

Searlova čínská komora



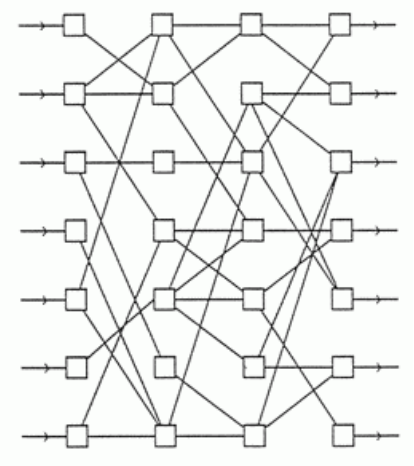
– *1980 John Searle*
– *argument proti mechanistické tezi*

– *modifikace Turingova testu:*
  – *v jedné místnosti stroj, který prošel čínskou variantou Turingova testu*
  – *v druhé místnosti John + přepis programu stroje (manuál „Jak mluvit čínsky)*
  – *oba zvenku dostávají otázky v čínštině, John vybírá odpovědi jen na základě manuálu (mechanicky)*

závěr: oba jsou stejně úspěšní, John nerozumí čínsky => stroj také nerozumí

– *námitky:* – *celek (John + manuál) může rozumět*
  – *John a procesor stroje nejsou vzájemně zaměnitelní*
  – *možnost vzniku emergentních jevů (porozumění)*

Konekcionismus



„Mentální stavy a procesy lze pojmout jako emergentní jevy dostatečně složitého dynamického systému.“

vlastnosti konekcionistických systémů:

1. *každý prvek může být v jednom z několika možných stavů aktivity*
2. *tento jeho stav závisí na stavu aktivity jiných prvků (případně i na vnějších stimulech)*
3. *závislostní vazby mají různé váhy, stupeň závislosti podle bodu 2 je určen vahami těchto vazeb*
4. *systém se vyvíjí, tj. váhy vazeb se v čase mění*

– *mozek je konekcionistický systém*
– *lze pozorovat analogie mezi mentálními jevy a chováním umělých konekcionistických systémů*