

# Silná umělá inteligence

4.4. 2006, Marek Kukačka

téma: [Umělá inteligence](#)

Při hledání tématu „umělá inteligence“ na Internetu či v knihovnách hrozí člověku velké zklamání. Velká většina informací, které tímto způsobem nalezne, totiž nepřipomíná nic, co by nějak souviselo s inteligencí, jak o ní mluvíme v souvislosti s lidským chováním. Témata, kterými se dnes obor Umělá Inteligence obvykle zabývá, jsou výsledkem snahy o napodobení různých aspektů lidského myšlení – tedy toho, jak si představujeme, že lidský mozek pracuje s některými typy informací, či jak řeší určité úlohy. Proto se dnes pod pojmem Umělá Inteligence skrývají věci jako neuronové sítě, expertní systémy, konečné automaty, algoritmy pro plánování a rozvrhování, zpracování přirozeného jazyka, obrazu nebo videa. Nic z toho při bližším prozkoumání nepřipomíná inteligentní počítače, kterými nás vytrvale zahrnují filmy hollywoodské produkce. V poslední době se ale začínají objevovat náznaky návratu ke kořenům, snahy o sestavení umělé mysli v počítači. Pojďme se podívat na dva zajímavé přístupy k tomuto problému.

Při hledání tématu „umělá inteligence“ na Internetu či v knihovnách hrozí člověku velké zklamání. Většina informací, které tímto způsobem nalezne, totiž nepřipomíná nic, co by nějak souviselo s inteligencí, jak o ní mluvíme v souvislosti s lidským chováním. Témata, kterými se dnes obor Umělá Inteligence obvykle zabývá, jsou výsledkem snahy o napodobení různých aspektů lidského myšlení – tedy toho, jak si představujeme, že lidský mozek pracuje s některými typy informací, či jak řeší určité úlohy. Proto se dnes pod pojmem Umělá Inteligence skrývají věci jako neuronové sítě, expertní systémy, konečné automaty, algoritmy pro plánování a rozvrhování, zpracování přirozeného jazyka, obrazu nebo videa. Nic z toho při bližším prozkoumání nepřipomíná inteligentní počítače, kterými nás vytrvale zahrnují filmy hollywoodské produkce.

Výzkum v oboru Umělé Inteligence se odchýlil od svého hlavního cíle. Tím je naučit počítač zvládat stejné problémy, jaké dnes umí řešit člověk. Zní to jednoduše, ovšem o to obtížnější tento úkol ve skutečnosti je. Lidský mozek je velice výkonný výpočetní stroj, jehož nasimulování v prostředí počítače je daleko za hranicemi možností dnešní technologie, a navíc je příliš komplexní, než aby ho bylo možné nahradit jedním či několika jednoduchými algoritmy.

Snaha o stejný přístup, jaký je praktikován ve fyzice, totiž aproximovat fungování složitého systému nějakým jednoduchým vzorečkem či principem, v případě umělé inteligence nevedla ke kýženým výsledkům. Výsledkem jsou však algoritmy použitelné v praxi, z nichž některé jsem jmenoval výše. Dnes jsou tyto algoritmy, inspirované způsobem lidského myšlení, úspěšně používány pro (alespoň přibližné) řešení určitých tříd problémů, pro které neexistují optimální postupy. Je možné, že v budoucnu se tyto algoritmy uplatní jako součásti rozsáhlejšího systému, který budeme moci nazvat umělou myslí. Hovořit tedy o úplném selhání GOFAI není zcela na místě.

Poslední dobou ale nalézám známky návratu k původním cílům umělé inteligence. Objevují se publikace a projekty (viz. [1] a [2]), jejichž témata a myšlenky se věnují konstrukci umělých inteligentních myslí se stejnými či lepšími schopnostmi, než mají lidské mozky.

Ačkoliv jsou výsledky těchto snah zatím pouze na úrovni teorií, některé z nich svým přístupem a pojetím slibují zajímavé budoucí přínosy.

## **Intelligence aneb Co se snažíme napodobit**

Inteligenci můžeme zhruba definovat jako schopnost vytvářet a manipulovat s vnitřními, mentálními modely okolního světa, díky čemuž můžeme předvídat následky vlastních akcí a jiných událostí a s ohledem na ně pak plánovat svou další činnost. U lidí je tato schopnost výsledkem komplexní spolupráce různých částí mozku.

Intelligence poskytla lidskému druhu jasnou evoluční výhodu. Je zajímavé ptát se, jaké kroky evolučního vývoje schopností mozku vedly od specializovaných reflexů u nižších živočichů až k obecné inteligenci. Jelikož evoluce upřednostňuje jedince s okamžitou výhodou, muselo k vývoji vyšších kognitivních funkcí dojít postupným zlepšováním schopností celého druhu. Možným vysvětlením je vývoj zdvojených systémů (tento postup evoluce využívá v mnoha případech – např. párové orgány), z nichž jeden později začal vykonávat nové, jedince zvýhodňující funkce. Mozek proto obsahuje některé evolučně staré systémy, například limbický systém, kolem kterých „vyrostly“ mladší části s novou funkcionalitou.

Při vývoji umělé intelligence budeme moci postupovat mnohem efektivněji, než jak postupovala evoluce při vývoji lidské mysli. Umělá mysl bude, na rozdíl od lidské, vytvořena s jasným cílem, což nám umožní optimalizovat její design. Je pravděpodobné, že díky optimalizacím bude možné provozovat UI na strojích s nižším výkonem, než jaký má lidský mozek. Tvůrci umělé intelligence by měli být schopni modelovat přímo složitější funkce lidského mozku, bez nutnosti simulovat práci jednotlivých neuronů, čímž dojde k dramatickému snížení hardwarových nároků. Funkcí, které bude třeba implementovat, bude stále velké množství, a nebudou nejjednodušší, stále to však bude méně náročné na výpočetní výkon než simulovat paralelní práci miliard či bilionů neuronů.

## **Architektury mysli**

Selhání „fyzikálního“ přístupu k vývoji umělé intelligence, tj. snahy o nalezení jednoho všemocného algoritmu obsahujícího „esenci intelligence“, donutilo výzkumníky v tomto oboru přiznat, že lidská mysl, a tedy i její simulace, se bude zřejmě skládat z většího počtu velmi odlišných funkcí. Přijetí představy mysli v podobě množství spolupracujících funkcí nám umožňuje podívat se na tuto konstrukci z různých nových úhlů pohledu.

Aaron Sloman, vedoucí projektu CogAff (Cognition and Affection, viz. [3]), zaměřuje svůj výzkum na definování a zkoumání architektur inteligentních agentů. Tyto architektury poskytují prostředky k charakterizaci procesů a stavů, vyskytujících se v myslích lidí, zvířat, či budoucích myslících počítačů. Ve článcích, uveřejněných v rámci projektu, popisuje komponenty těchto architektur, zavádí metody pro porovnávání rozdílů mezi různými typy architektur, a také například rozebírá původ a přínos výskytu emocí v inteligentních systémech.

Zajímavé je zohlednění evolučního vývoje při popisu lidské intelligence pomocí Slomanových architektur. Sloman popisuje architekturu lidské mysli jako trojvrstvou. První, evolučně nejstarší, je takzvaná reaktivní vrstva. Jsou v ní uloženy funkce mapující smyslové vjemy přímo na určité akce, bez jakéhokoli mechanismu plánování či zvažování alternativ. Tyto funkce jsou napevno „zadrátované“ do řídicího systému agenta. V umělém inteligentním

systému by funkce v této vrstvě mohly být implementovány například pomocí systému IF-THEN pravidel. Nevýhodou tohoto designu je malá možnost adaptace na změnu vnějších podmínek, možná pouze pomocí přirozeného výběru přes několik generací. Výhodou je naopak rychlost reakcí tohoto systému řízení. Agenti vybaveni čistě reaktivním řídicím systémem obsazují ve svém ekosystému niky, které těmto výhodám a nevýhodám odpovídají – je možné, že takovýto čistě reaktivním typem agenta je hmyz.

Další vrstvou je tzv. deliberativní vrstva. Ta funguje nad reaktivní vrstvou a umožňuje přizpůsobit chování agenta při změně vnějšího světa i v případě, že reaktivní vrstva nemá pro tento nový stav zakódovanou reakci. Funkce v deliberativní vrstvě jsou schopny vybrat akci nejvhodnější pro daný stav vnějšího světa, a to pomocí otestování možných variant na mentálním modelu. Takovéto plánování budoucích akcí vyžaduje přítomnost nějaké formy krátkodobé paměti, pro uložení momentálně probíraných možností. Právě možnost uložení dočasných plánů pro vyhodnocení je podle Slomana hlavní rozdíl mezi deliberativní a čistě reaktivní architekturou. Pro efektivní plánování je také třeba znát požadovaný výsledek akcí agenta, proto architektura s deliberativní vrstvou musí obsahovat systém motivátorů a filtrů pozornosti s proměnlivým prahem (o těchto konstrukcích více později).

Třetí vrstvou je tzv. vrstva meta-řízení (v originále meta-management). Jejím úkolem je sledování a řízení dějů v deliberativní vrstvě. Je zde sledováno a zaznamenáváno, jaká rozhodnutí byla přijata, které plány selhaly a proč, nebo s jakými problémy se funkce v deliberativní vrstvě potýkaly. Tyto údaje jsou pak vyhodnocovány podle hlavních cílů a úkolů agenta, a podle výsledků jsou ovlivňovány děje v nižších vrstvách. Typickými funkcemi tohoto systému by byl výběr efektivnějšího řešení problému známého z dřívějšíka, řešení konfliktu mezi cíli, či změna strategie při selhání příliš mnoha úkolů.

Kolem vrstvy meta-řízení je mnoho nedorozumění otázek. Kupříkladu pro různé úrovně řešení problémů by možná bylo potřeba mít navíc vrstvu meta-meta-řízení atd., to je však možné řešit tím, že vrstvě meta-řízení umožníme (rekurzivně) sledovat samu sebe. Další problém přináší otázka co přesně má tato vrstva sledovat – ve složitém systému nelze monitorovat vše, nehledě na to, že sledování sebe sama by vyústilo v nekonečnou smyčku. Sloman navrhuje řešit tento problém přidáním „introspektivních“ mechanismů, které z nízkoúrovňových informací v reaktivní vrstvě dokáží abstrakci a zjednodušením vyrobit vhodná data pro deliberativní a meta-řídicí vrstvu.

Sloman se ve svých pracích ve velké míře věnuje emocím a jejich v architekturách agentů. Emoce jsou podle něj emergentní jev, nastávající při změně pozornosti agenta. Pozornost je v architekturách přítomna v podobě filtru, který určuje, jaký cíl (či jaké cíle) agenta se uplatní při plánování akcí. Filtr má proměnlivý práh a propouští pouze cíle, jejichž důležitost tento práh překročí. Pokud změna v prostředí způsobí náhlý přechod cílů přes filtr, dojde k přeplánování, a právě tyto změny v systému jsou podle Slomana příčiny emočních stavů. Sloman dělí emoce podle toho, kterou vrstvu architektury ovlivňují. Primární emoce, působící na reaktivní vrstvu, jsou kupříkladu strach či překvapení. Sekundární emoce, jako potěšení z úspěchu či zklamání při neúspěchu, jsou reakce na plnění cílů v deliberativní vrstvě. Terciární emoce ovlivňují vrstvu meta-řízení a souvisí s psychickými stavy. Sloman mezi ně řadí například lásku či smutek způsobený ztrátou blízké osoby.

## Úrovně organizace inteligentní mysli

Další práce, zabývající se konstrukcí silné UI, je pojednání o úrovních organizace od Eliezera Yudkowského (viz [4]). Zaměřuje se na struktury, které by měly tvořit umělou mysl na různých úrovních, podobně jako je lidské tělo sestaveno z atomů, z nichž se skládají molekuly, ze kterých se skládají buňky, atd. Procesy, jichž se tyto struktury účastní, nastiňuje jen velmi zhruba. Při rozebírání jednotlivých úrovní se snaží řešit problémy, které se vyskytly ve stejných oblastech při minulých neúspěšných pokusech o sestavení fungující UI.

Yudkowski popisuje UI realizovanou v prostředí počítače na pěti úrovních organizace. Nejnižší je úroveň zdrojového kódu, na které se nacházejí funkce, datové struktury a další programovací konstrukce. Tato úroveň je zcela v režii programátora, její funkcionalita je jím pevně určena. Jelikož analogií v lidském mozku jsou neurony a jejich propojení a fungování, je zde hlavním problémem implementace takového systému, který dokáže podporovat vyšší úrovně organizace v prostředí počítače stejně, jako to dělají neuronové sítě v mozku. Vzhledem k rozdílnosti počítače a mozku, obzvláště co do výpočetního výkonu, je tento problém velmi obtížně řešitelný.

Další úroveň je vrstva smyslových oblastí. Na této úrovni probíhá analýza vstupních informací ze smyslů, podobně jako probíhá zpracování vizuálních vjemů v lidské sítnici a zrakovém centru v mozkové kůře. Tato vrstva je opět vytvořena programátorem, k jejímu návrhu je však třeba dostatečně porozumět roli, kterou hrají smysly v lidském mozku. Je pravděpodobné, že kromě zpracování vstupních dat se smyslové oblasti účastní také procesu myšlení, například poskytováním prostoru pro vytváření mentálních obrazů. Při vytváření těchto smyslových oblastí (autor navrhuje například oblast pro vnímání zdrojového kódu programů) je tedy třeba uplatnit znalosti o jejich úloze v lidském mozku. Důležitá je také vhodná volba algoritmů pro analýzu informací. Již na sítnici se ve vrstvách neuronů provádí detekce hran a pohybu – jaké informace ale bude potřeba vybrat z bloku zdrojového kódu?

Nad smyslovými oblastmi se nachází úroveň konceptů. Slovem „koncept“ zde autor označuje vzor abstrahovaný ze zkušenosti, pravidelnost ve výsledcích analýzy vstupních dat ve smyslové oblasti. Konceptem může být například „červená“ nebo „vypadat jako cihla“, v podstatě jsou koncepty základní stavební bloky struktur na další úrovni organizace, které jsou ekvivalentní lidským myšlenkám. Aby byl systém konceptů dostatečně robustní, měl by být zcela v režii UI (která se jednotlivé koncepty učí ze svých zkušeností), nikoliv v rukou programátora, který by je ručně vyráběl.

V oblasti symbolické UI byly koncepty reprezentovány holými atomy jazyka LISP, případně byly charakterizovány svým umístěním v sémantických sítích. Yudkowski přisuzuje omezenost symbolické UI právě tomu, že byly (ve snaze modelovat jen vyšší kognitivní funkce) ignorovány nižší úrovně organizace, na kterých koncepty stojí, čímž tyto koncepty ztratily veškerou sémantiku (resp. celý jejich význam byl založen na jejich pojmenování – symbol pojmenovaný *jablko* reprezentoval jablko, a podobně).

Myšlenky, další z úrovní organizace UI, jsou okamžité kombinace konceptů. Určitým způsobem ovlivňují vznik mentálních obrazů, například myšlenka obsahující koncept „strom“ může zapříčinit vytvoření představy stromu ve vizuální sensorické oblasti. Samy jsou také reakcí na vytvořené představy, čímž se vytváří cyklus interakce, ve kterém jsou vytvářeny další myšlenky jako reakce na ty předchozí.

Poslední úroveň organizace Yudkowski označuje jako „uvažování“ (deliberation), podle jeho slov se tím vyhýbá použití zprofanovaného slova „vědomí“. Tato úroveň zahrnuje činnosti myslí implementované sekvencemi myšlenek, tedy věci, které mysl dělá, spíše než ty, které ji tvoří. Jsou zde zahrnuty činnosti jako plánování akcí za účelem splnění cíle, předvídání chování okolního prostředí, vysvětlení a porozumění okolním jevům a jejich vazbám, a další. Autor v souvislosti s touto vrstvou rozebírá otázky introspekce a rozpoznání svého „já“ umělou inteligencí.

Jak autor zdůrazňuje průběžně v celé práci, považuje jím popsanou vnitřní organizaci UI za nutnou, ale nikoliv postačující podmínku pro funkční implementaci. Představený návrh úrovní organizace tvoří rámec, ke kterému je třeba vybrat ty správné algoritmy, datové struktury a další součásti, aby celý systém fungoval podle očekávání.

## Shrnutí

Ve své práci jsem nastínil dva rozdílné přístupy ke zkoumání silné umělé inteligence, kvůli jejich odlišnosti mi ale přijde obtížné provést nějaké hlubší porovnání. Výsledky projektu CogAff Aarona Slomana by se v budoucnu mohly stát měřítkem pro klasifikaci architektur inteligentních agentů. Poskytují také velice zajímavý pohled na funkce, které by inteligentní agent měl implementovat. Je však obtížné si představit, že by se na základě těchto teorií dala sestavit konkrétní implementace. Na to je projekt CogAff příliš vágní, ovšem být konkrétní v detailech nikdy nebylo jeho cílem. Práce Eliezera Yudkowského na druhé straně popisuje představu autora o strukturách uvnitř umělé myslí mnohem detailněji. I zde se autor vyhýbá rozebírání bližších detailů, podle mě se tím ale jen rozumně brání přílišnému zjednodušení problému. Tato práce se mi zdá být mnohem použitelnějším teoretickým základem pro konkrétní implementaci. Dalo by se kupříkladu vyjít z jednoduchého modelu, splňujícího požadavky na základní úroveň organizace, a postupně tento model obohacovat. Také je mi sympatický autorův přístup. Yudkowsky svou teorii vypracoval s ohledem na příčiny selhání předchozích pokusů o silnou UI, přičemž se snaží, podle mého názoru úspěšně, těmto problémům vyhnout.

Netroufám si hodnotit přínos těchto prací pro budoucí realizaci silné UI, těší mě ale skutečnost, že výzkum na tomto poli nestagne (jak jsem si ještě nedávno myslel) a že je v něm vidět snaha o nové pohledy na problematiku a poučení se z předchozích nezdarů. Na další vývoj silné UI se proto dívám optimisticky a jsem velmi zvědav na knihu [1], která má vyjít v dubnu 2006 a v níž bude, podle předběžného obsahu, shrnuto množství prací ze současného výzkumu v této oblasti.

## Odkazy a reference

[1] Artificial General Intelligence, Ben Goertzel a Cassio Pennachin, nakl. Springer, zatím nepublikováno (původně plánováno na únor 2006, pak na duben, nyní 15. května – a pořád nic!)

<http://www.springer.com/sgw/cda/frontpage/0,,4-147-22-43950079-0,00.html>

[2] Artificial General Intelligence Research Institute

<http://agiri.org/>

[3] The Cognition and Affection Project

<http://www.cs.bham.ac.uk/~axs/cogaff/>

[4] Levels of Organization in General Intelligence, Eliezer Yudkowsky, Singularity Institute for Artificial Intelligence

<http://www.singinst.org/LOGI/>

Marek Kukačka

Autor je studentem Matematicko-fyzikální fakulty UK.  
Zaměřuje se na umělou inteligenci.