هوش مصنوعی پیشرفته

فصل ۲۷

حال و آینده‌ی هوش مصنوعی

Artificial Intelligence: Present and Future

کاظم فولادی
دانشکده مهندسی برق و کامپیوتر
دانشگاه تهران

http://courses.fouladi.ir/aai

هوش مصنوعی

حال و آینده‌ی هوش مصنوعی

۱

مقدمه

مسئله‌ی طراحی هوش مصنوعی

AI Design Problem

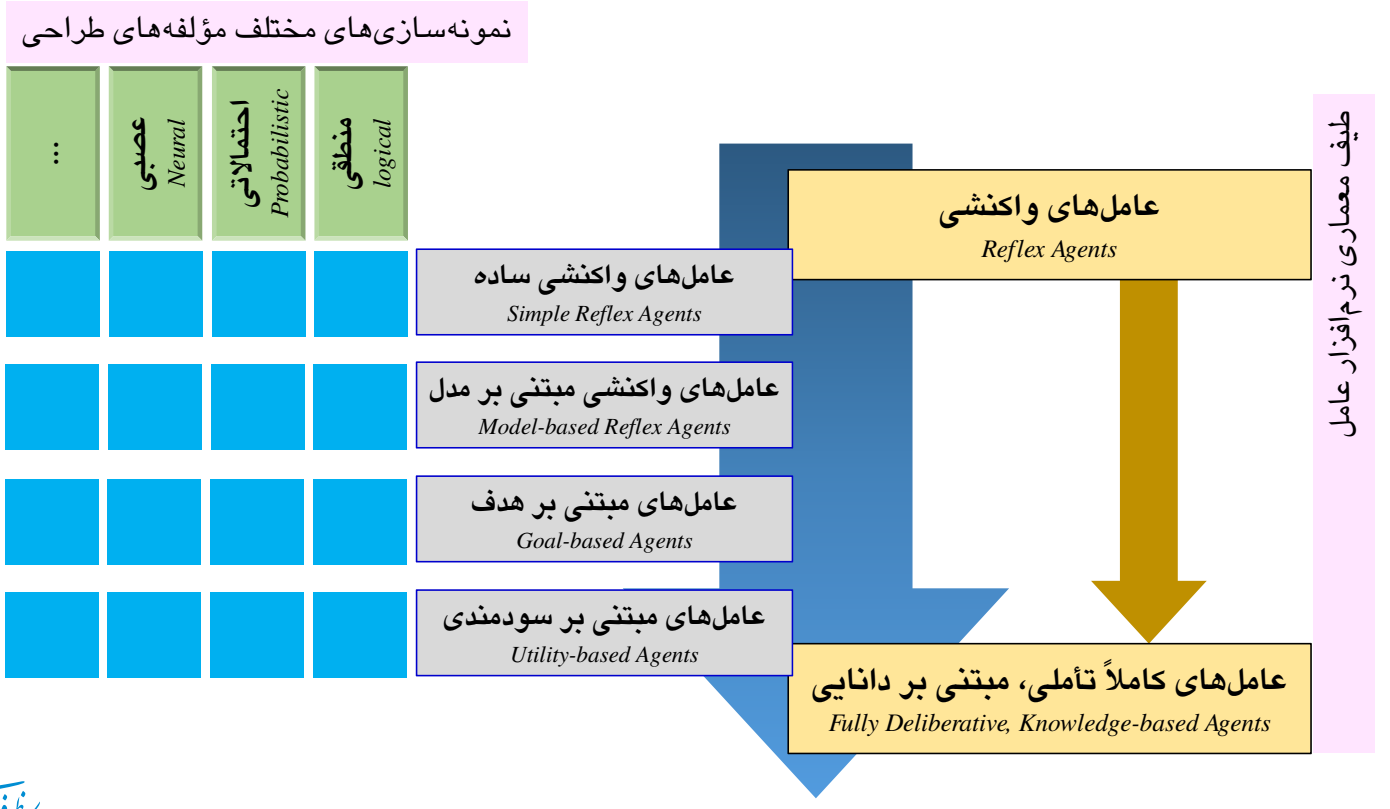
نگاه یکپارچه به هوش مصنوعی در قالب طراحی عامل رسیونال

مسئله‌ی طراحی عامل هوشمند *Intelligent Agent Design Problem*			
وابسته است به:			
ادراک‌ها *Percepts*	**کنش‌ها** *Actions*	**سودمندی** *Utility*	**محیط** *Environment*
ادراک‌ها و کنش‌های موجود برای عامل		تابع سودمندی که رفتار عامل باید آن را ارضا کند.	طبیعت محیط

مسئله‌ی طراحی هوش مصنوعی

طراحی‌های گوناگون برای عامل‌ها

AI Design Problem



مسئله‌ی طراحی هوش مصنوعی

AI DESIGN PROBLEM

برای همه‌ی طراحی‌ها و مؤلفه‌های عامل،
پیشرفت‌های شگرفی رخ داده است:
هم در درک علمی ما و هم در قابلیت‌های تکنولوژیک ما

اما پرسشی که وجود دارد:

آیا همه‌ی این پیشرفت‌ها به یک عامل هوشمند همه‌منظوره که می‌تواند در گستره‌ی وسیعی از محیط‌ها به‌خوبی عمل کند، منجر می‌شود؟

حال و آینده‌ی هوش مصنوعی

AI: Present and Future

مؤلفه‌های یک عامل هوشمند

برای ارزیابی دانسته‌ها و نادانسته‌های ما

معماری‌های عامل

کدام معماری برای عامل هوشمند؟

آیا در جهت درستی پیش می‌رویم؟

هدف طراحی عامل رسیونال هدف درستی است؟

نتیجه‌گیری

چه می‌شود اگر هوش مصنوعی موفق شود؟

هوش مصنوعی

حال و آینده‌ی هوش مصنوعی

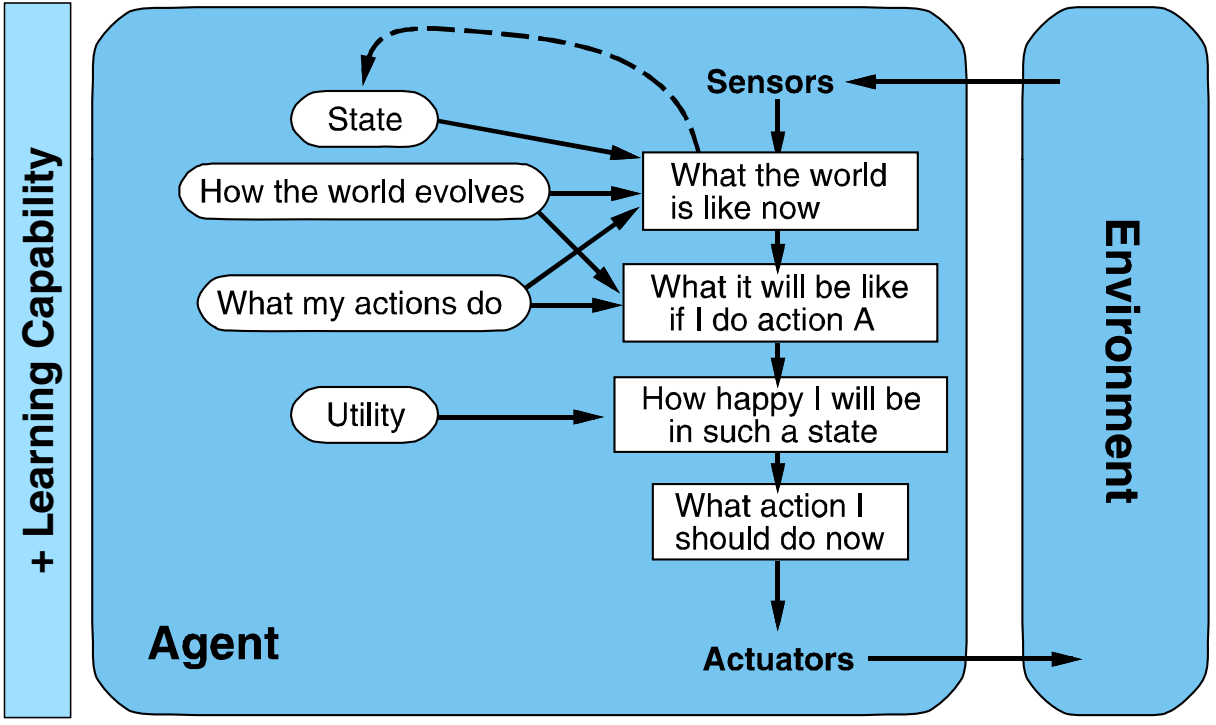
۲

مؤلفه‌های عامل

مؤلفه‌های عامل

در عامل مبتنی بر سودمندی (+ مبتنی بر مدل)

AGENT COMPONENTS

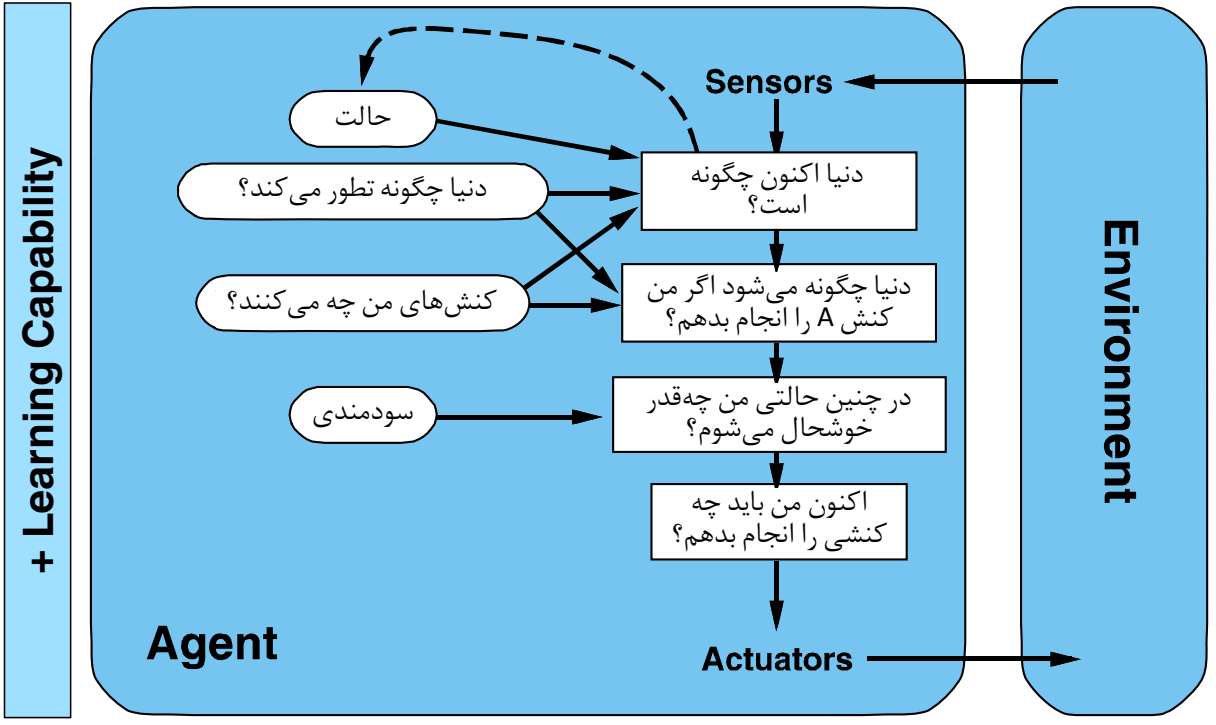


عامل مبتنی بر سودمندی به‌عنوان کلی‌ترین شکل طراحی عامل

مؤلفه‌های عامل

در عامل مبتنی بر سودمندی (+ مبتنی بر مدل)

AGENT COMPONENTS

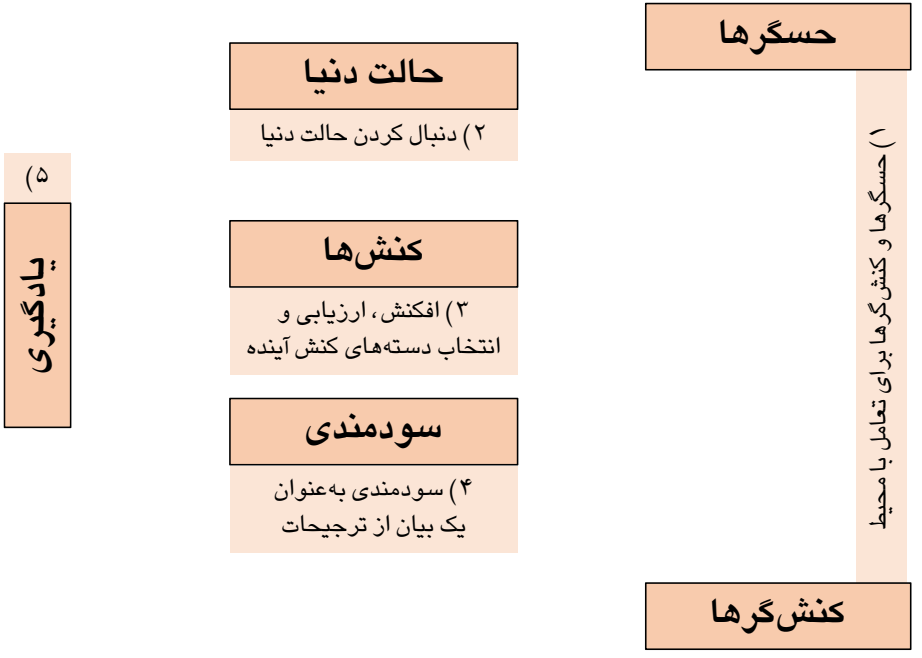


عامل مبتنی بر سودمندی به‌عنوان کلی‌ترین شکل طراحی عامل

مؤلفه‌های عامل

در عامل مبتنی بر سودمندی (+ مبتنی بر مدل)

AGENT COMPONENTS



مؤلفه‌های عامل

۱) حسگرها و کنش‌گرها برای تعامل و محیط

1) Interaction with the Environment through Sensors and Actuators

در بخش بزرگی از تاریخچه‌ی هوش مصنوعی،
موضوع تعامل با محیط از طریق حسگرها و کنش‌گرها،
یک نقطه ضعف به‌شمار می‌آید.

سیستم‌های هوش مصنوعی، اغلب به‌گونه‌ای ساخته شده‌اند که
انسان‌ها باید به آنها ورودی بدهند و خروجی‌های آنها را تفسیر کنند.

سیستم‌های رباتیک، بر روی وظیفه‌های سطح پایینی تمرکز کرده بودند که در آنها
استدلال و طرح‌ریزی سطح بالا بسیار غایب است.
علت، هزینه‌ی بالا و تلاش مهندسی لازم برای ساخت ربات‌های واقعی کارآمد است.
اما این وضعیت در سال‌های اخیر با ظهور ربات‌های قابل برنامه‌ریزی آماده‌به‌کار تغییر کرده است:
استفاده از دوربین‌های CCD با کیفیت بالا، ارزان و کوچک، محرکه‌های موتوری فشرده،
تکنولوژی MEMS (سیستم‌های میکروالکترومکانیک): ژیروسکوپ، شتاب‌سنج کوچک، ...

⇓

مدت‌هاست که برای محیطهای فیزیکی سیستم‌های هوش مصنوعی، دیگر بهانه‌ای وجود ندارد
+ وجود اینترنت به‌عنوان یک محیط کاملاً جدید.

مؤلفه‌های عامل

۲) دنبال کردن حالت دنیا

2) KEEPING TRACK OF THE STATE OF THE WORLD

یکی از قابلیت‌های اصلی لازم برای یک عامل هوشمند،
دنبال کردن حالت دنیا است.

ملزومات دنبال کردن حالت دنیا	
ادراک *Perception*	**به‌هنگام‌سازی بازنمایی‌های داخلی** *Updating of Internal Representations*

ابزارها		
تکنیک‌های منطق گزاره‌ای	**تکنیک‌های منطق مرتبه اول**	**الگوریتم‌های فیلترینگ**

برای دنبال کردن محیط‌های نامطمئن

مسئله‌ی عدم اطمینان در هویت
Problem of Identity Uncertainty

در مورد اشیا در محیط‌های نامطمئن

«نمی‌دانیم کدام شیئ کدام است؟»
این مسئله در AI مبتنی بر منطق به‌شدت نادیده گرفته شده است و به‌طور کلی فرض شده است که ادراکات نمادهایی را وارد می‌کنند که اشیا را شناسایی می‌کنند.

مؤلفه‌های عامل

۳) افکنش، ارزیابی و انتخاب دسته‌های کنش آینده

3) PROJECTING, EVALUATING, AND SELECTING FUTURE COURSES OF ACTION

اعمال ساختار سلسله‌مراتبی بر روی رفتار عامل
برای چیرگی بر گام‌های متعدد یک کنش
(مثلاً هزاران یا میلیون‌ها کنش ابتدایی لازم برای انجام یک کنش در عامل واقعی)

ابزارها		
الگوریتم‌های طرح‌ریزی	**الگوریتم‌های تصمیم‌گیری**	**الگوریتم‌های یادگیری**
با بازنمایی سلسله‌مراتبی و منطق مرتبه اول	تصمیم‌گیری در شرایط عدم اطمینان بر اساس مفاهیم جستجو مبتنی بر حالت	یادگیری‌های سلسله‌مراتبی، مانند یادگیری تقویتی سلسله‌مراتبی

مؤلفه‌های عامل

۴) سودمندی به‌عنوان یک بیان از ترجیحات

4) UTILITY AS AN EXPRESSION OF PREFERENCES

نیاز به توابع سودمندی واقع‌گرایانه وجود دارد.
Realistic

قرار دادن تصمیم‌های رسیونال بر اصل ماکزیمم‌سازی متوسط سودمندی خیلی کلی است
و از بسیاری مشکلات ناشی از روی‌کردهای کاملاً مبتنی بر هدف
(تداخل اهداف / دستاوردهای غیر مطمئن) جلوگیری می‌کند.

برای ایجاد توابع سودمندی واقع‌گرایانه، کار اندکی صورت گرفته است.
ثابت شده است که تجزیه‌ی ترجیحات حالت‌های پیچیده دشوار است:
یک دلیل آن می‌تواند این باشد که ترجیحات حالت‌ها ازتاریخچه‌ی ترجیحات حالت‌ها به‌دست می آید
که توسط توابع پاداش (reward) توصیف می‌شوند:
حتی اگر توابع پاداش ساده باشند، تابع سودمندی متناظر می‌تواند بسیار پیچیده باشد.

پیشنهاد: در تعریف توابع پاداش باید جدی عمل کرد به‌گونه‌ای که عامل متوجه شود از آن انتظار چه کاری می‌رود.

مؤلفه‌های عامل

۵) یادگیری

5) LEARNING

رویکردهای یادگیری استقرائی		
بانظارت	**بدون نظارت**	**تقویتی**

یادگیری ماشینی، پیشرفت بسیار اندکی در مورد مسئله‌ی بسیار مهم
ساخت بازنمایی‌های جدید در سطوح انتزاع بالاتر از لغت‌نامه‌ی ورودی
داشته است.

مثلاً: عامل چگونه می‌تواند محمول‌هایی مانند *Office* و *Cafe* را تولید کند، اگر توسط انسان به او گفته نشود؟

مثلاً: درک از یک کنش سطح بالاتر که متشکل از چند کنش ابتدائی است؟ (ایده‌هایی مثل **شبکه‌های باور عمیق**)

در مقابل، برای یک لغت‌نامه‌ی مشخص از ویژگی‌ها و مفاهیم،
توسعه‌ی روش‌های آماری و منطقی بسیار قدرت‌مند برای
بازشناسی پیش‌بینانه‌ی الگوها موفق بوده است (حتی بیش از توان انسان).

بیشتر روش‌های یادگیری، بازنمایی «فاکتورگیری شده» را فرض کرده‌اند. (نیاز به تعمیم به بازنمایی ساختاری)

میزان موفقیت و فهم عامل با پیچیدگی بازنمایی کاهش می‌یابد.

استفاده از دانایی پس‌زمینه ضعیف است.

هوش مصنوعی

حال و آینده‌ی هوش مصنوعی

۳

معماری‌های عامل

معماری‌های عامل

AGENT ARCHITECTURES

کدام‌یک از معماری‌های عامل باید برای یک عامل هوشمند استفاده شود؟

معماری‌های عامل

AGENT ARCHITECTURES

کدام‌یک از معماری‌های عامل باید برای یک عامل هوشمند استفاده شود؟

پاسخ: همه‌ی آنها!!

عامل‌های واکنشی ساده
Simple Reflex Agents

عامل‌های واکنشی مبتنی بر مدل
Model-based Reflex Agents

عامل‌های مبتنی بر هدف
Goal-based Agents

عامل‌های مبتنی بر سودمندی
Utility-based Agents

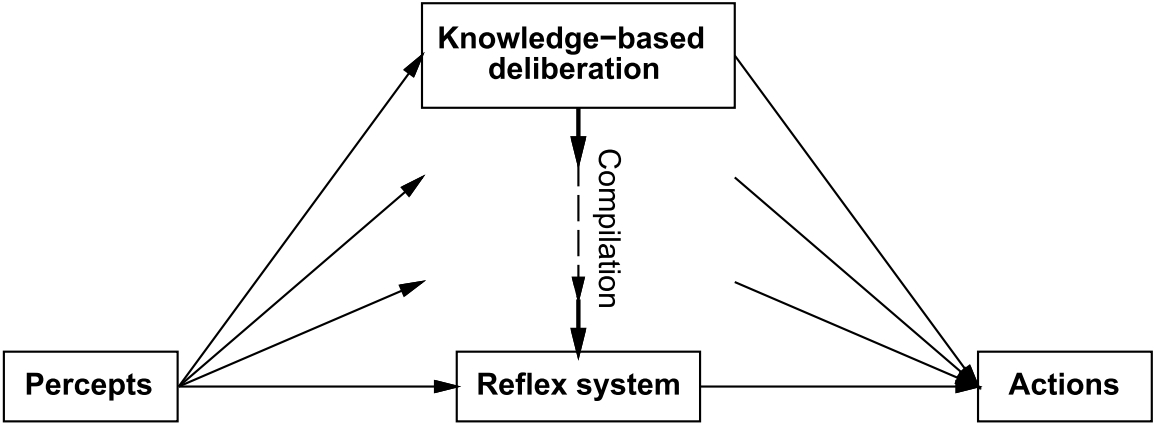
معماری واکنشی برای وضعیت‌هایی که زمان اصل است لازم است.
تأملات مبتنی بر دانایی به عامل امکان طرح‌ریزی پیشاپیش می‌دهد.

⇓

یک عامل کامل به هر دو نیاز دارد: استفاده از معماری هیبرید

معماری هیبرید

HYBRID ARCHITECTURE



مرزهای میان مؤلفه‌های مختلف تصمیم‌گیری ثابت نیست.

کامپایل کردن
Compilation

- تبدیل تصمیم‌گیری تأملی به مکانیسم‌های کارآمدتر واکنشی
- تعمیم دادن تأمل به واکنش

استفاده از این الگو در معماری‌هایی چون SOAR و THEO

هوش مصنوعی بی‌درنگ

REAL-TIME AI

عامل‌ها باید راهی برای کنترل تأمل‌هایشان داشته باشند:
عامل‌ها باید بتوانند در زمان نیاز به کنش، تأمل را متوقف کرده و از زمان باقیمانده برای تأمل در اجرای مفیدترین محاسبات استفاده کنند.

مثلاً: تصمیم عامل تاکسی هوشمند، به محض دیدن تصادف در مقابلش:
باید در کمتر از یک ثانیه تصمیم به ترمز یا تغییر مسیر بگیرد.

⇐ نیاز به:

هوش مصنوعی بی‌درنگ
Real-Time AI

با حرکت سیستم‌های هوش مصنوعی به سمت دامنه‌های پیچیده‌تر،
همه‌ی مسائل به مسائل بی‌درنگ تبدیل می‌شوند؛
زیرا عامل هرگز زمان کافی برای حل دقیق مسائل تصمیم‌گیری نخواهد داشت.

هوش مصنوعی بی‌درنگ

الگوریتم‌های مناسب برای وضعیت‌های تصمیم‌گیری عمومی‌تر

REAL-TIME AI

تصمیم‌گیری بی‌درنگ	
الگوریتم‌های هرزمانی *Anytime Algorithms*	**فرا–استدلال نظریه تصمیمی** *Decision-Theoretic Meta-reasoning*

هوش مصنوعی بی‌درنگ

الگوریتم‌های مناسب برای وضعیت‌های تصمیم‌گیری عمومی‌تر: الگوریتم‌های هرزمانی

REAL-TIME AI

تصمیم‌گیری بی‌درنگ

فرا-استدلال نظریه تصمیمی
Decision-Theoretic Meta-reasoning

الگوریتم‌های هرزمانی
Anytime Algorithms

- راه‌حل با گذر زمان به‌تدریج بهبود می‌یابد،
- اما همیشه یک راه‌حل آماده وجود دارد.

مثل روش: عمیق‌کننده‌ی تکراری در بازی

الگوریتمی که کیفیت راه‌حل آن به‌تدریج در طول زمان بهبود می‌یابد به‌گونه‌ای که هرگاه ناگهانی متوقف شود، همیشه یک تصمیم مستدل آماده دارد.
مثال ساده: مثل IDS در انجام بازی

هوش مصنوعی بی‌درنگ

الگوریتم‌های مناسب برای وضعیت‌های تصمیم‌گیری عمومی‌تر: فرا–استدلال نظریه تصمیمی

REAL-TIME AI

تصمیم‌گیری بی‌درنگ

فرا–استدلال نظریه تصمیمی
Decision-Theoretic Meta-reasoning

الگوریتم‌های هرزمانی
Anytime Algorithms

- تصمیم می‌گیرد که کدام محاسبه باید انجام شود.
- بر اساس بده‌بستان میان هزینه و فایده

این روش از نظریه‌ی ارزش اطلاعات برای انتخاب محاسبات استفاده می‌کند.
ارزش محاسبه تابعی از هزینه و فایده‌ی کنش.
فرا–استدلال یکی از جنبه‌های یک معماری عمومی بازتابی است: یک معماری که تأمل در مورد موجودیت‌های محاسباتی و کنش‌هایی که درون خود معماری رخ می‌دهد را ممکن می‌کند.
(با تعریف یک فضای حالت مشترک از حالت محیط و حالت محاسباتی خود عامل)

هوش مصنوعی

حال و آینده‌ی هوش مصنوعی

۴

آیا در جهت درستی پیش می‌رویم؟

آیا در جهت درستی پیش می‌رویم؟

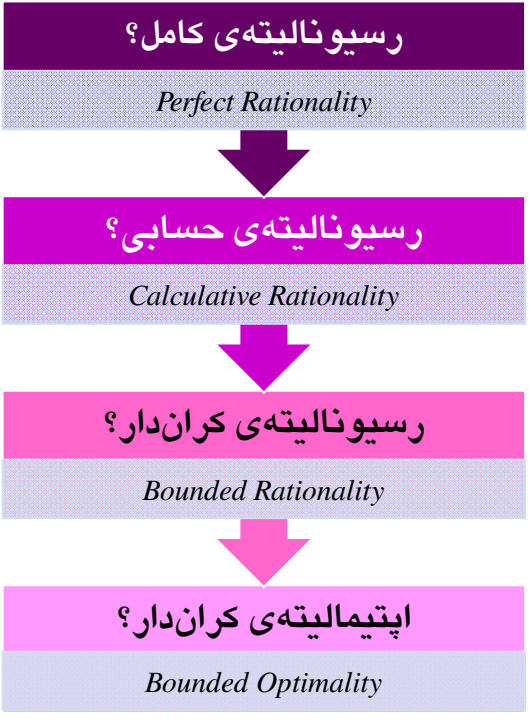
ARE WE GOING IN THE RIGHT DIRECTION?

مسیر فعلی هوش مصنوعی،
بیشتر شبیه بالا رفتن از درخت است یا سفر موشکی برای رسیدن به ماه؟

هدف ما، ساختن عامل‌هایی است که به‌صورت رسیونال کنش می‌کنند.
اما ...
دست‌یابی به رسیونالیته‌ی کامل ـ یعنی همیشه کار درست انجام دادن ـ
در محیط‌های پیچیده ممکن نیست.
نیازمندی‌های محاسباتی بسیار بالا هستند.
با این وجود، فرضیه‌ی رسیونالیته‌ی کامل نقطه‌ی شروع خوبی برای بحث است.

هدف هوش مصنوعی دقیقاً چیست؟

WHAT EXACTLY THE GOAL OF AI IS?



رسیونالیته‌ی کامل

PERFECT RATIONALITY

یک عامل رسیونال کامل، در هر لحظه از زمان به‌گونه‌ای کنش می‌کند که
امید سودمندی آن ماکزیمم شود
با در اختیار داشتن اطلاعاتی که از محیط به دست آورده است.

- نظریه‌ی تصمیم کلاسیک
- ماکزیمم‌سازی امید سودمندی (سودمندی مورد انتظار)

این روی‌کرد برای محیطهای پیچیده امکان‌ناپذیر است!

رسیونالیته‌ی حسابی

CALCULATIVE RATIONALITY

رسیونالیته‌ی حسابی، مفهومی از رسیونالیته است
که به طور ضمنی در طراحی عامل‌های منطقی و نظریه تصمیمی استفاده می‌شود.

یک عامل رسیونال حسابی، نهایتاً آنچه در ابتدای فرآیند تأمل آن،
گزینه‌ی رسیونال بوده است را برمی‌گرداند.

اما در بیشتر محیطها، جواب درست در زمان نادرست ارزشی ندارد

عامل رسیونال حسابی نهایتاً یک گزینه‌ی رسیونال تولید می‌کند که:
- این گزینه در ابتدای تأمل عامل معتبر بوده است.
- با مقداری محاسبه تعریف می‌شود.
- ممکن است زمان زیادی طول بکشد.

رسیونالیته‌ی کران‌دار

BOUNDED RATIONALITY

هربرت سایمون (۱۹۵۷م) مفهوم رسیونالیته‌ی کامل (یا حتی تقریباً کامل) را با مفهوم رسیونالیته‌ی کران‌دار جایگزین کرد.

رسیونالیته‌ی کران‌دار اصولاً با ارضا کردن (satisficing) کار می‌کند:
زمان تأمل تنها آن‌قدر کافی است که به جواب به‌اندازه‌ی کافی خوب برسیم.

سایمون برای این کار جایزه‌ی نوبل در اقتصاد را برد!
به نظر او این مدل مفیدی برای رفتارهای انسان در بسیاری از حالت‌هاست.

این نظریه
- به دلیل عدم تعریف «به‌اندازه‌ی کافی خوب» برای عامل‌های هوشمند مناسب نیست.
- ارضاکنندگی فقط یکی از روش‌های کار با منابع محدود است! نه همه‌ی آنها.

اپتیمالیته‌ی کران‌دار

BOUNDED OPTIMALITY

عامل بهینه‌ی کران‌دار بر اساس منابع محاسباتی‌اش، تا حد ممکن بهینه رفتار می‌کند.

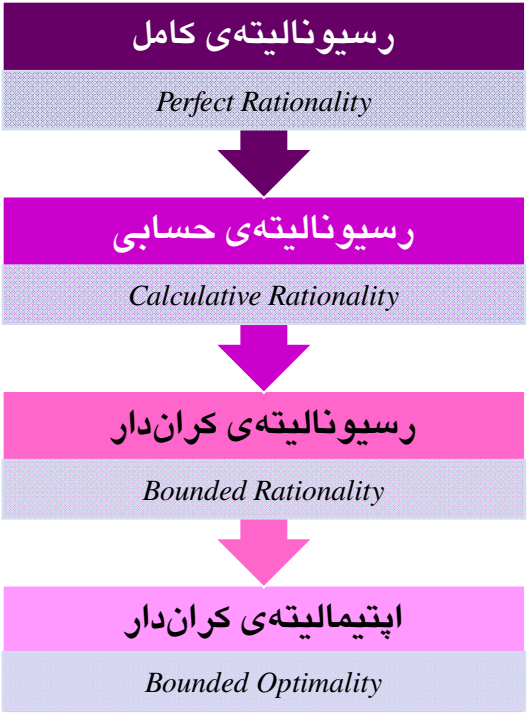
متوسط سودمندی برنامه‌ی عامل برای عامل اپتیمال کران‌دار
حداقل به‌اندازه‌ی ماکزیمم متوسط سودمندی هر برنامه‌ی عامل دیگری است
که روی ماشین مشابهی اجرا شده است.

از میان چهار امکان مطرح شده، به‌نظر می‌رسد اپتیمالیته‌ی کران‌دار امیدواری بیشتری
برای تهیه‌ی یک مبنای نظری برای هوش مصنوعی ایجاد می‌کند.

- امکان‌پذیرترین گزینه برای رسیونالیته‌ی عامل هوشمند.
  - عامل اپتیمال کران‌دار: دسترس‌پذیر (بر خلاف رسیونال کامل)
  - عامل اپتیمال کران‌دار: کاربردی در دنیای واقعی (بر خلاف رسیونال حسابی)
  - عامل اپتیمال کران‌دار: کاربردی در دنیای واقعی (بر خلاف رسیونال کران‌دار)
- مزیت **ممکن بودن دست‌یابی:** همیشه حداقل یک بهترین برنامه وجود دارد.

نکته: بهینگی کران‌دار، متوجه برنامه‌های عامل بهینه است نه کنش‌های بهینه.

سیر عقب‌نشینی در فلسفه‌ی رسیونالیته



هوش مصنوعی

حال و آینده‌ی هوش مصنوعی

۵

چه می‌شود اگر هوش مصنوعی موفق شود؟

چه می‌شود اگر هوش مصنوعی موفق شود؟

WHAT IF AI DOES SUCCEED?

می‌توانیم انتظار داشته باشیم موفقیت‌های سطح متوسط در هوش مصنوعی
در زندگی روزمره‌ی همه‌ی افراد بشر تأثیر بگذارد.

قابلیت‌های تکنولوژیکی در این سطح، در توسعه‌ی سلاح‌های خودمختار نیز به‌کار گرفته می‌شود!

به نظر می‌رسد که موفقیت‌های سطح بالا در هوش مصنوعی،
زندگی عمده‌ی نوع بشر را تغییر می‌دهد.
در این سطح، سیستم‌های هوش مصنوعی تهدید مستقیم‌تری برای خودمختاری، آزادی و حتی بقای انسان است!

⇓

نمی‌توانیم پژوهش‌های هوش مصنوعی را از پی‌آمدهای اخلاقی آن جدا کنیم.

نتیجه

CONCLUSION

می‌بینیم که هوش مصنوعی پیشرفت بزرگی در تاریخچه‌ی کوتاهش داشته است،
اما جمله‌ی پایانی آلن تورینگ در مقاله‌ی
ماشین‌های محاسبه و هوشمندی (۱۹۵۰م)
امروزه هنوز معتبر است:

We can see only a short distance ahead,
but we can see that much remains to be done.

تنها مسافت کوتاهی از پیش رو را می‌توانیم ببینیم،
اما می‌توانیم ببینیم چیزهای زیادی باقی مانده تا انجام شود.

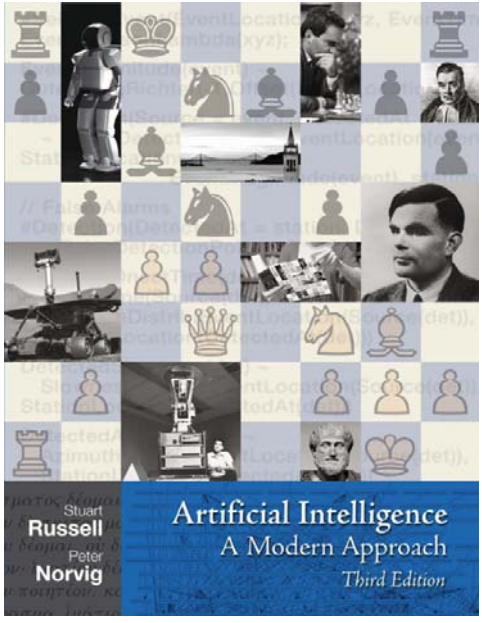
هوش مصنوعی

حال و آینده‌ی هوش مصنوعی

۶

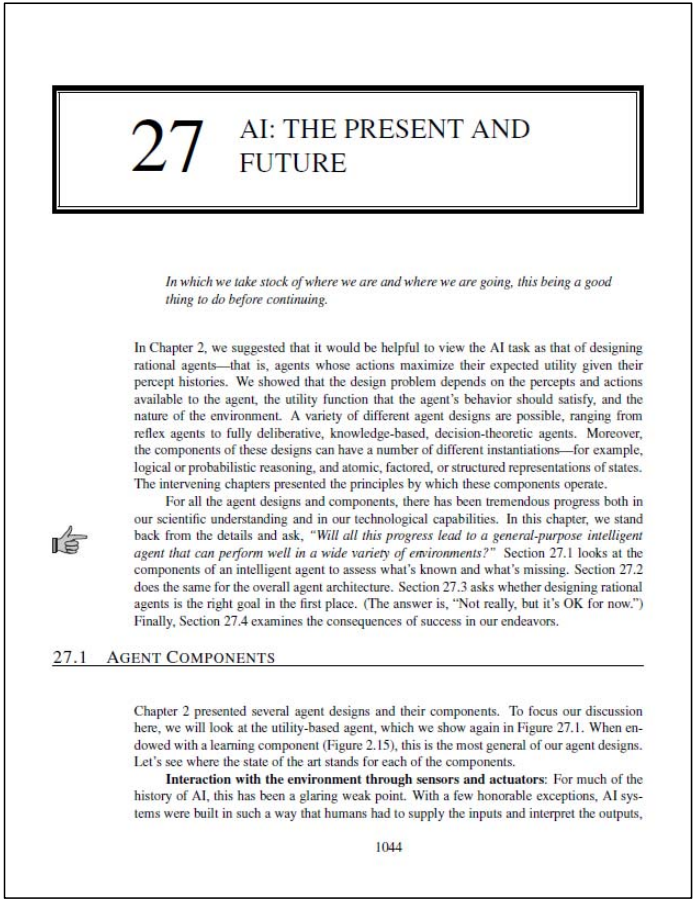
منابع

منبع اصلی



Stuart Russell and Peter Norvig,
Artificial Intelligence: A Modern Approach,
3rd Edition, Prentice Hall, 2010.

Chapter 27



27 AI: THE PRESENT AND FUTURE

In which we take stock of where we are and where we are going, this being a good thing to do before continuing.

In Chapter 2, we suggested that it would be helpful to view the AI task as that of designing rational agents—that is, agents whose actions maximize their expected utility given their percept histories. We showed that the design problem depends on the percepts and actions available to the agent, the utility function that the agent's behavior should satisfy, and the nature of the environment. A variety of different agent designs are possible, ranging from reflex agents to fully deliberative, knowledge-based, decision-theoretic agents. Moreover, the components of these designs can have a number of different instantiations—for example, logical or probabilistic reasoning, and atomic, factored, or structured representations of states. The intervening chapters presented the principles by which these components operate.

For all the agent designs and components, there has been tremendous progress both in our scientific understanding and in our technological capabilities. In this chapter, we stand back from the details and ask, *"Will all this progress lead to a general-purpose intelligent agent that can perform well in a wide variety of environments?"* Section 27.1 looks at the components of an intelligent agent to assess what's known and what's missing. Section 27.2 does the same for the overall agent architecture. Section 27.3 asks whether designing rational agents is the right goal in the first place. (The answer is, "Not really, but it's OK for now.") Finally, Section 27.4 examines the consequences of success in our endeavors.

27.1 AGENT COMPONENTS

Chapter 2 presented several agent designs and their components. To focus our discussion here, we will look at the utility-based agent, which we show again in Figure 27.1. When endowed with a learning component (Figure 2.15), this is the most general of our agent designs. Let's see where the state of the art stands for each of the components.

Interaction with the environment through sensors and actuators: For much of the history of AI, this has been a glaring weak point. With a few honorable exceptions, AI systems were built in such a way that humans had to supply the inputs and interpret the outputs,

1044