

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



سیستم‌های چند عاملی

درس ۱

مروری بر

هوش مصنوعی و عامل‌های هوشمند

A Review on
Artificial Intelligence (AI) and Intelligent Agents (IA)

کاظم فولادی قلعه

دانشکده مهندسی، پردیس فارابی

دانشگاه تهران

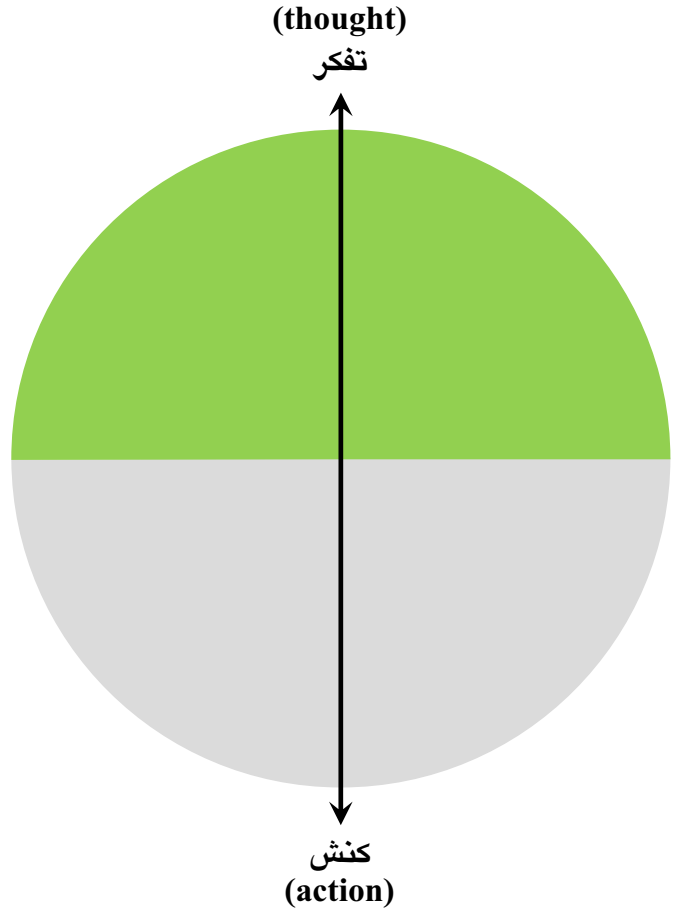
<http://courses.fouladi.ir/mas>

هوش مصنوعی و عامل های هوشمند

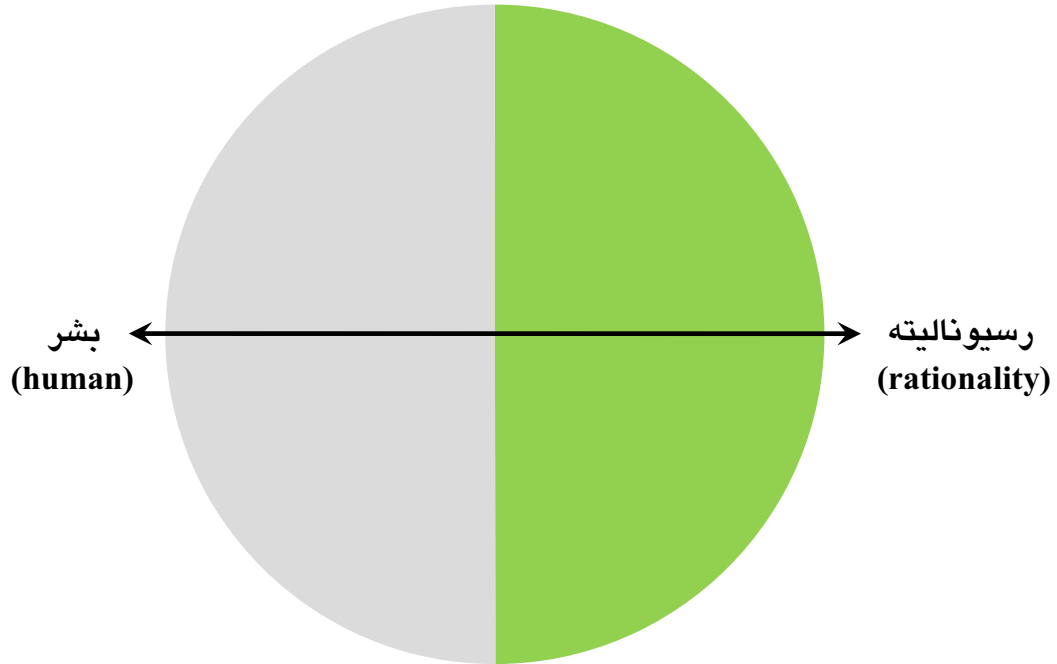
۱

هوش مصنوعی (AI)

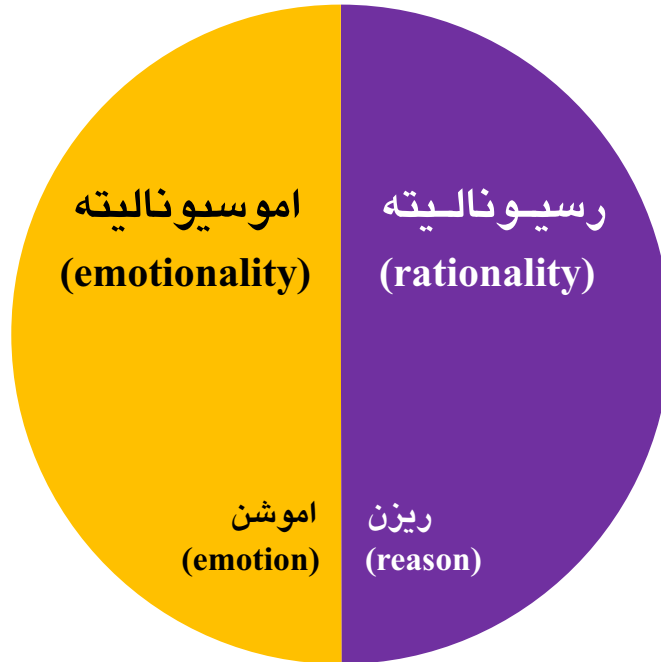
هوش: در تفکر یا کنش



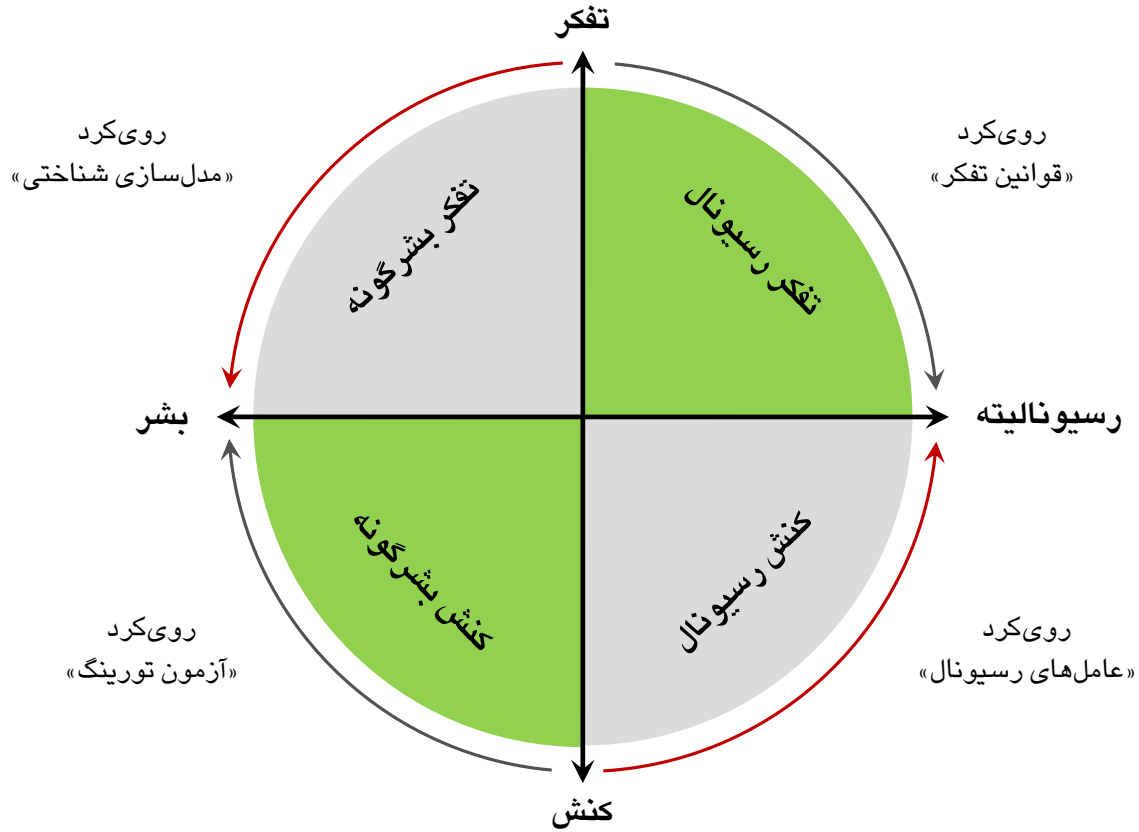
هوش: ایده‌آل هوشمندی (رسیونالیته) یا بشر



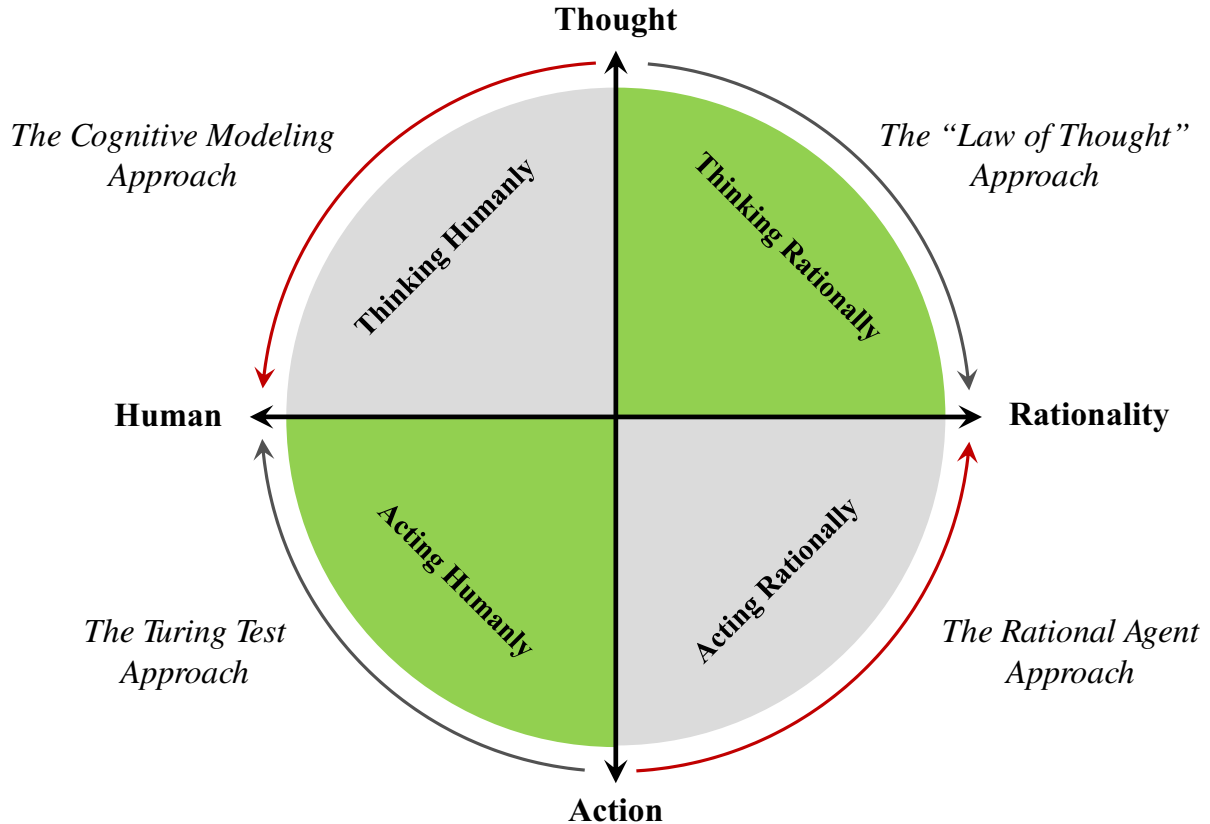
ساحات بشر در سایکولوژی



روی‌کردهای تعریف هوش مصنوعی

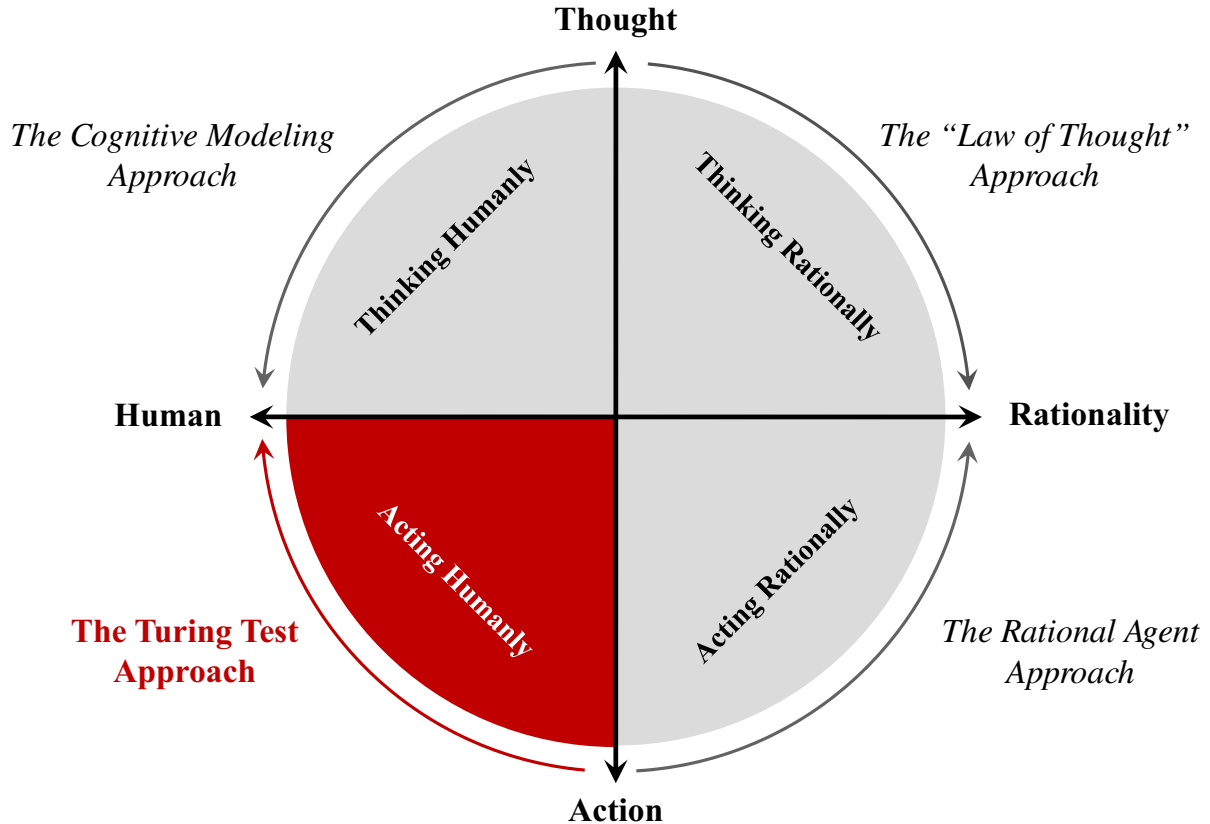


رویکردهای تعریف هوش مصنوعی



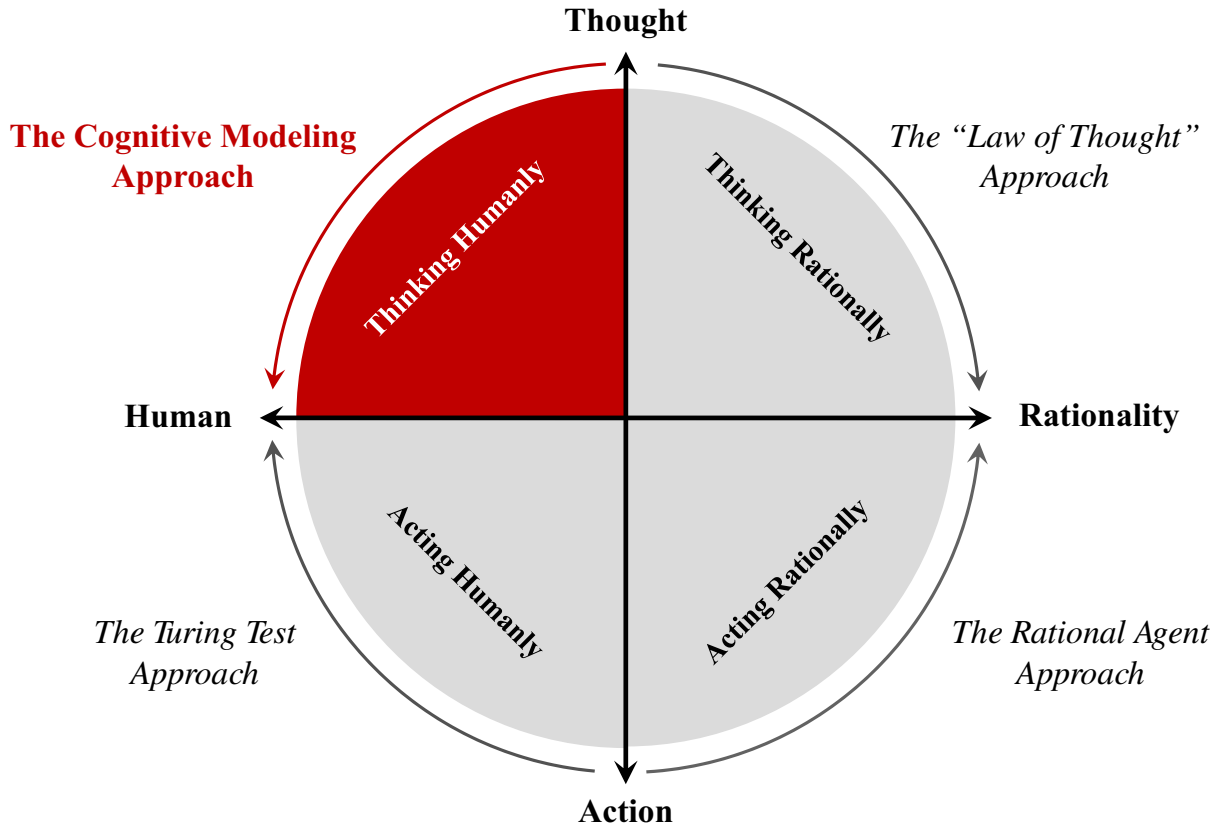
رویکردهای تعریف هوش مصنوعی

(۱) رویکرد «آزمون تورینگ» / کنش بشرگونه



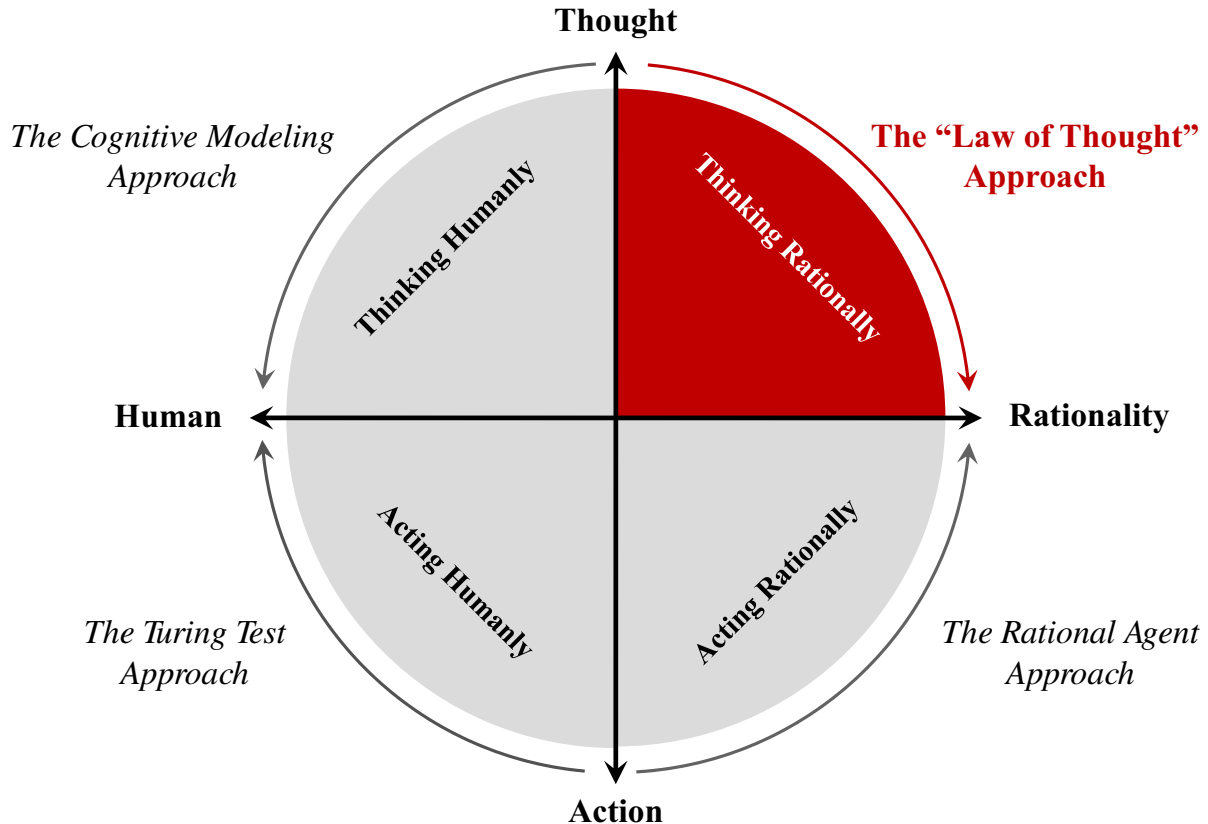
رویکردهای تعریف هوش مصنوعی

(۲) رویکرد «مدلسازی شناختی» / تفکر بشرگونه



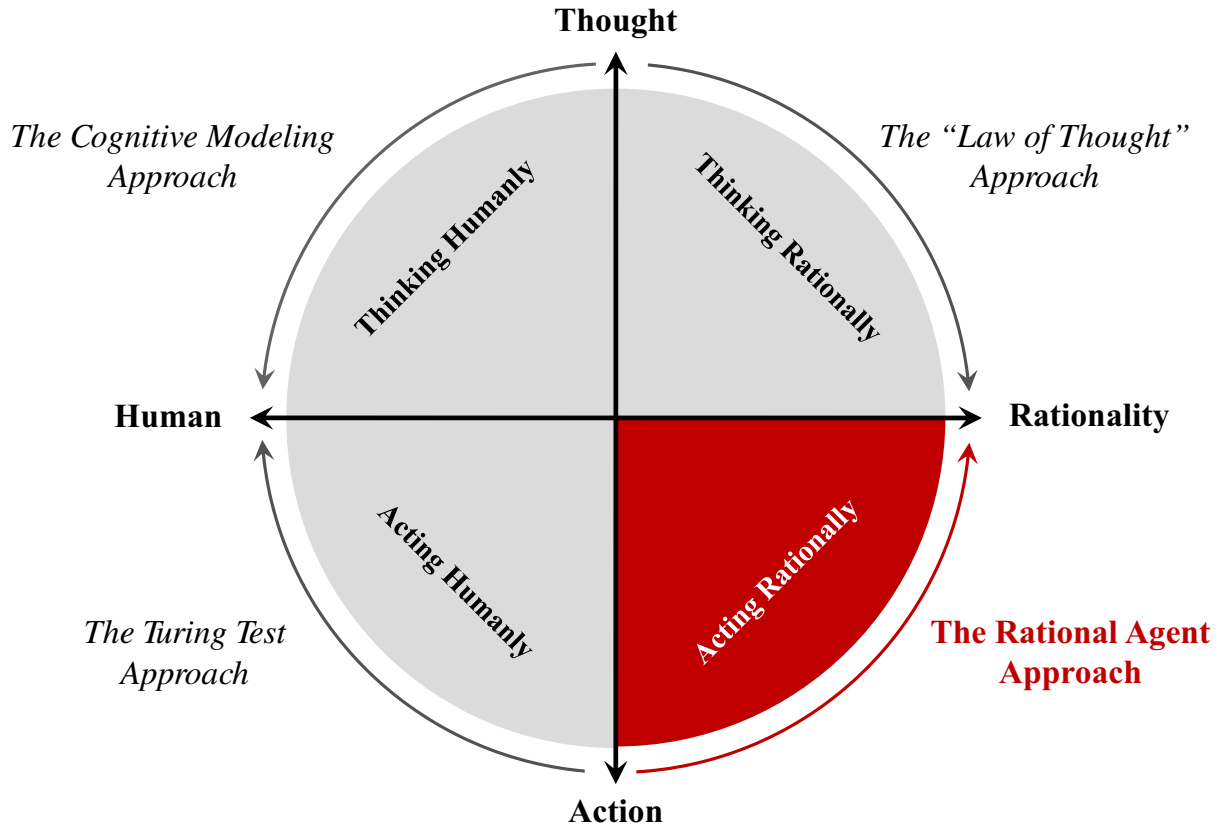
رویکردهای تعریف هوش مصنوعی

(۳) رویکرد «قوانین تفکر» / تفکر رسیونال



رویکردهای تعریف هوش مصنوعی

(۴) رویکرد «عامل‌های رسیونال» / کنش رسیونال



عامل

AGENT

عامل:
کننده‌ی کار
(چیزی که کنش می‌کند)

مشخصه‌های عامل‌های کامپیوتری (در مقابل برنامه‌های کامپیوتری معمولی)

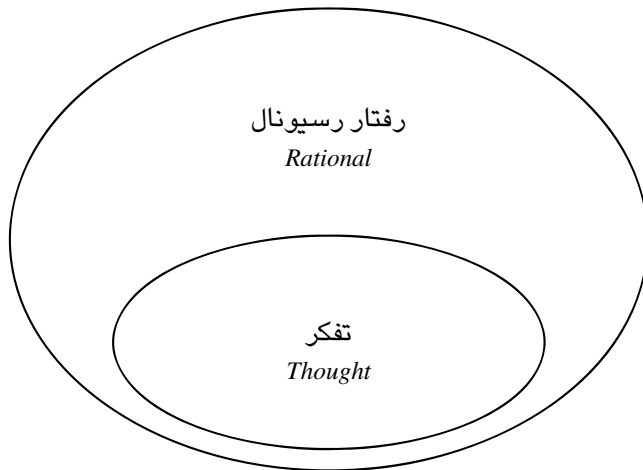
...	ایجاد و پیگیری اهداف <i>create and pursue goals</i>	وفق‌یابی با تغییر <i>adapt to change</i>	استمرار در زمان طولانی <i>persist over a prolonged time</i>	درک محیط <i>perceive the environment</i>	عملکرد خودمختار <i>operate autonomously</i>
-----	---	--	---	--	---

عامل رسیونال

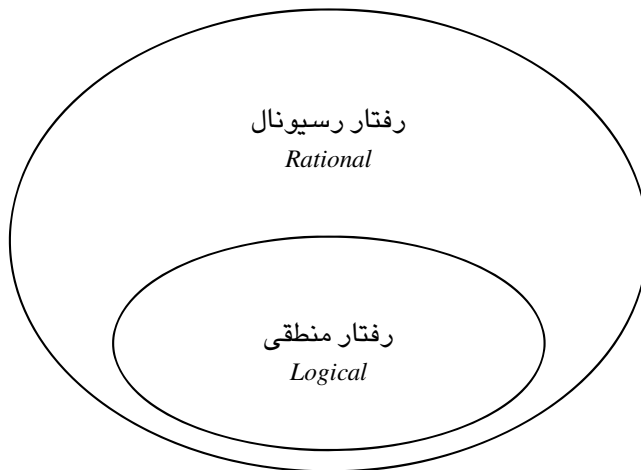
RATIONAL AGENT

عامل رسیونال (کننده‌ی کار خوب)	
در شرایط عدم اطمینان	در شرایط اطمینان
انجام کنش با بهترین برآمد مورد انتظار	انجام کنش با بهترین برآمد

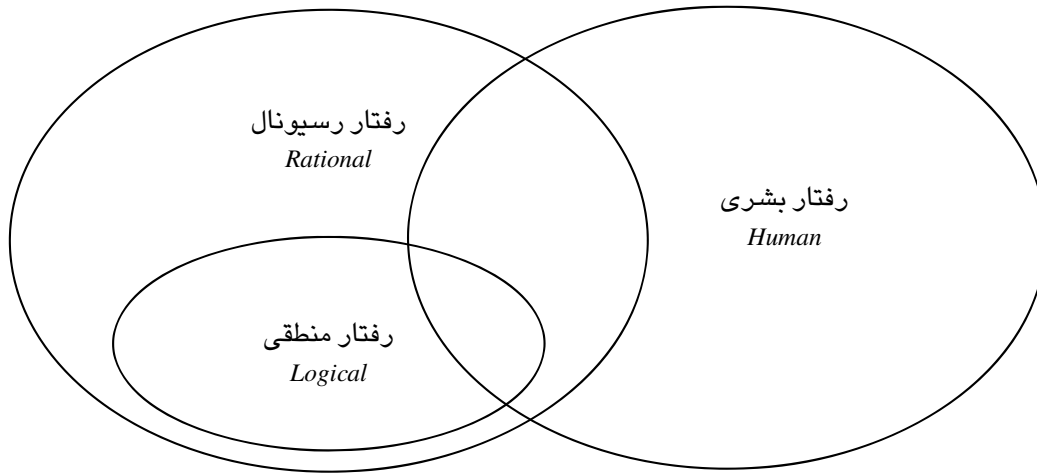
نسبت رفتار رسیونال با تفکر



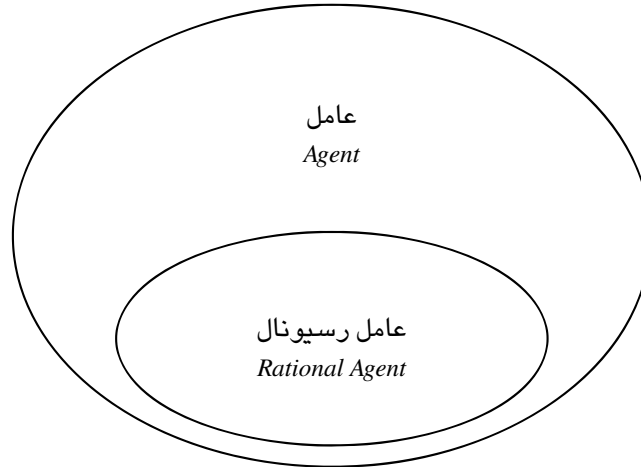
نسبت رفتار رسیونال با رفتار منطقی



نسبت رفتار رسیونال و رفتار منطقی با رفتار بشری



عامل رسیونال



عامل: موجودیتی که درک می‌کند و کنش انجام می‌دهد.

موضوع هوش مصنوعی: طراحی عامل‌های رسیونال

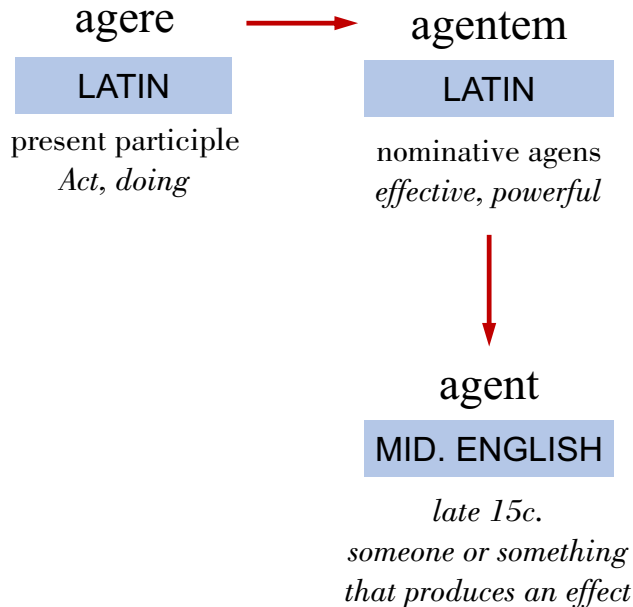
سیستم های چند عاملی

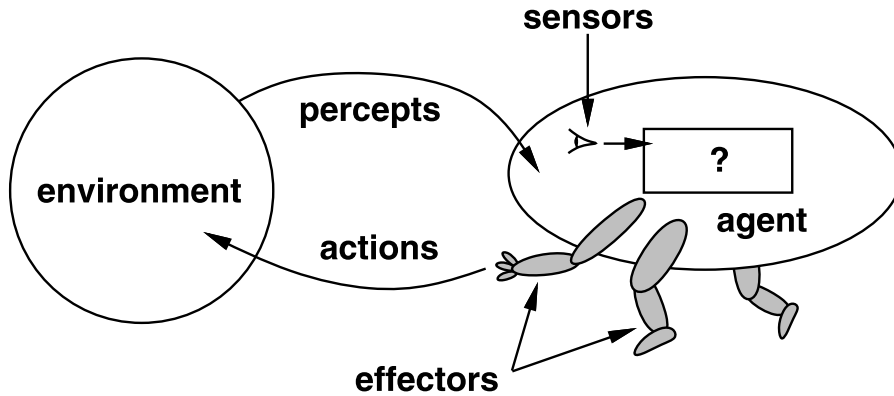
هوش مصنوعی و عامل های هوشمند

۲

عامل

اتیمولوژی: Agent



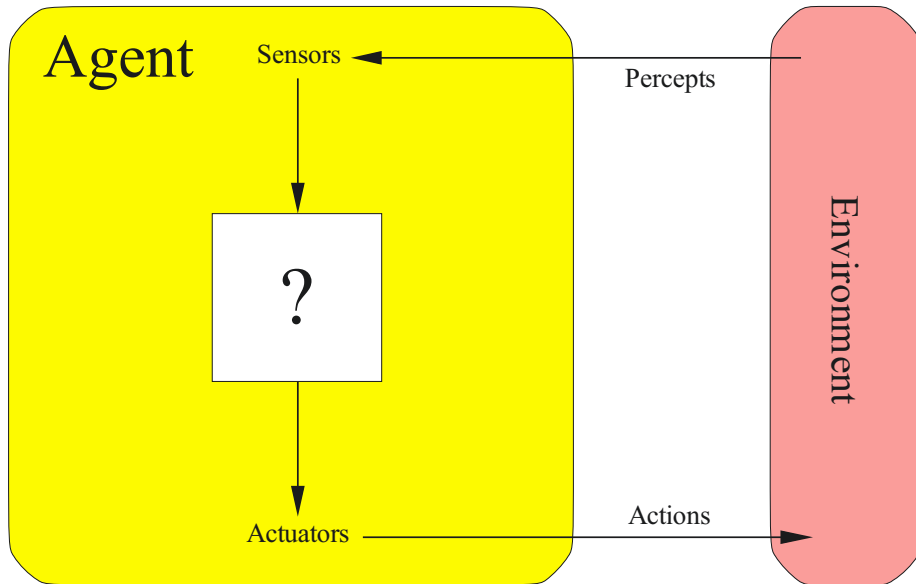
AGENT

عامل موجودیتی است که در **محیط** کار انجام می‌دهد.

عامل

AGENT

عامل چیزی است که محیط را از طریق حسگرهای خود درک می‌کند و سپس از طریق کنش‌گرهای خود روی آن کنش انجام می‌دهد.



عامل

مثال

حواس پنج‌گانه: چشم، گوش و ...	حسگر	عامل انسانی <i>Human agent</i>
دست و پا، زبان و ...	کنش‌گر	
دوربین، فاصله‌یاب مادون قرمز و ...	حسگر	عامل روبات <i>Robotic agent</i>
موتور و ...	کنش‌گر	
دنباله‌ای از بیت‌ها	حسگر	عامل نرم‌افزاری <i>Software agent</i>
دنباله‌ای از بیت‌های کد شده	کنش‌گر	

مفهوم **عامل** وسیله‌ای برای مطالعه‌ی هوش مصنوعی است.

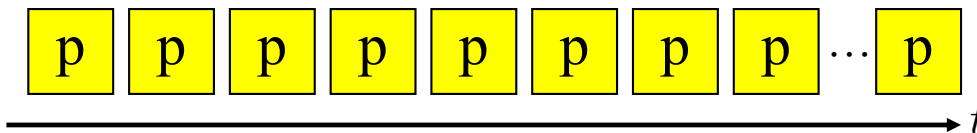
تقسیم‌بندی اشیا به «عامل» و «غیرعامل» نادرست است.
بر اساس نگاه طراح هر چیزی می‌تواند یک «عامل» باشد.

ادراک و دنباله‌ی ادراکی

ادراک (percept)، ورودی‌های ادراکی عامل در هر لحظه‌ی داده شده است.

p

دنباله‌ی ادراکی (percept sequence)، تاریخچه‌ی کامل همه‌ی چیزهایی است که عامل درک کرده است.

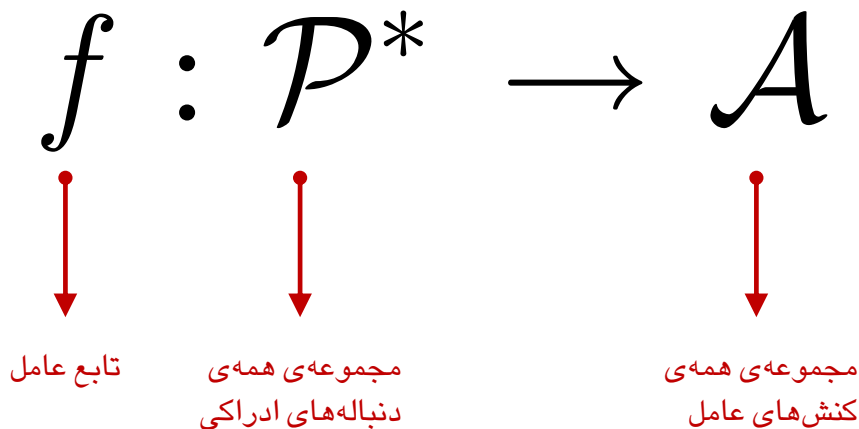


انتخاب کنش در هر لحظه توسط عامل،
می‌تواند به کل دنباله‌ی ادراکی مشاهده شده تا آن لحظه وابسته باشد،
اما به هیچ چیزی که تا آن لحظه آن را درک نکرده است، وابسته نیست.

تابع عامل

AGENT FUNCTION

رفتار عامل، با **تابع عامل** توصیف می‌شود که هر دنباله‌ی ادراکی داده شده را به یک کنش نگاشت می‌دهد.



برنامه‌ی عامل

AGENT PROGRAM

برنامه‌ی عامل، تابع عامل را بر روی یک معماری فیزیکی پیاده‌سازی و اجرا می‌کند.



نسبت تابع عامل با برنامه‌ی عامل

پیاده‌سازی می‌کند

تابع عامل <i>Agent Function</i>	برنامه‌ی عامل <i>Agent Program</i>
انتزاعی - مجرد (abstract)	انضمامی - مجسم (concrete)
ریاضی	برنامه‌نویسی
ورودی: یک دنباله‌ی ادراکی	ورودی: ادراک لحظه‌ای
-	قابل اجرا روی یک معماری
ممکن یک تابع عامل، برنامه‌ی عامل نداشته باشد.	یک تابع عامل، می‌تواند چند برنامه‌ی عامل داشته باشد.

تعیین عامل‌ها

Must

کدام موجودیت **می‌باید**
به‌عنوان عامل دیده شود؟

May

کدام موجودیت **می‌تواند**
به‌عنوان عامل دیده شود؟

پاسخ: براساس تاثیر عامل‌ها بر معیار کارآیی

هوش مصنوعی و عامل های هوشمند

۳

عامل
هوشمند
(IA)

عامل رسیونال

RATIONAL AGENT

عامل رسیونال:

عاملی که رفتار **خوب** انجام می‌دهد.



تقریب اول: رفتار خوب؛
رفتاری که باعث **بیشترین موفقیت** عامل شود.



نیاز به روشی برای **اندازه‌گیری موفقیت**
(**معیار کارآیی**)

معیار کارایی

معیار خوب بودن

PERFORMANCE MEASURE

معیار کارایی

میزان موفقیت رفتار یک عامل را نشان می‌دهد.

«معیار ارزیابی با در نظر گرفتن پی‌آمدهای رفتار آن عامل»

هر عامل در یک محیط بر اساس ادراک‌هایی که دریافت می‌کند، یک دنباله از کنش‌ها تولید می‌کند؛ دنباله‌ی کنش‌های عامل، موجب دنباله‌ای از حالت‌های محیط می‌شود؛ اگر این دنباله‌ی حالت‌های محیط، **مطلوب** بود، آن‌گاه عامل به خوبی عمل کرده است.

↓
مفهوم مطلوب بودن در معیار کارایی احصا می‌شود:
معیار کارایی هر دنباله از حالت‌های محیط را ارزیابی می‌کند.

تذکر: در معیار کارایی، حالت‌های محیط مهم است، نه حالت‌های عامل
یعنی معیار کارایی را بر اساس آنچه در محیط می‌خواهیم طراحی می‌کنیم،
نه بر اساس آنچه فکر می‌کنیم عامل باید انجام بدهد.

معیار کارآیی

ملاحظات

مقایسه‌ی معیار کارآیی عینی و ذهنی

معیار کارآیی عینی <i>Objective Performance Measure</i>	معیار کارآیی ذهنی <i>Subjective Performance Measure</i>
نظر محیط برای ارزیابی عامل (توسط طراح و سازنده‌ی عامل)	نظر شخصی خود عامل برای ارزیابی او
ثابت: غیر قابل تغییر توسط عامل	متغیر: قابل تغییر توسط خود عامل
معیار کارآیی بر اساس هدف انتخاب می‌شود	معیار کارآیی با نیت موفق‌تر شدن عامل انتخاب می‌شود
مزیت: نیازی به پرسش از خود عامل نیست	مشکل: ممکن است عامل نتواند پاسخ بدهد
مزیت: عامل نمی‌تواند بسادگی محیط را فریب بدهد.	مشکل: ممکن است عامل پاسخ گمراه‌کننده و اشتباه بدهد

- نباید عامل را به دلیل یک چیز غیر قابل درک جریمه کند.
- نباید عامل را به دلیل یک کنش غیر قابل انجام جریمه کند.
- نباید عامل را به دلیل یک پی‌آمد غیر قابل پیش‌بینی جریمه کند.
- باید ارزیابی عامل در طول مدت کار (+ اول + آخر کار) باشد.
- باید هر دوی فرآیند و نتیجه‌ی کار عامل را در نظر بگیرد.

مشخصات یک معیار کارآیی مناسب

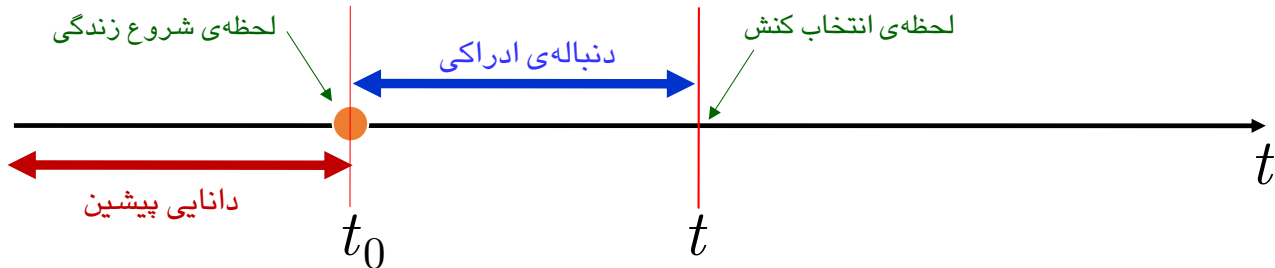
یک معیار کارآیی ثابت برای همه‌ی وظیفه‌ها و همه‌ی عامل‌ها وجود ندارد:
طراح باید متناسب با شرایط، معیار کارآیی را تدبیر کند.

رسیونالیته

پارامترهای وابسته

RATIONALITY

رسیونالیته در هر زمان وابسته به چهار مورد است:			
۱	۲	۳	۴
معیار کارایی	دانایی پیشینی	کنش‌ها	دنباله‌ی ادراکی
<i>Performance Measure</i>	<i>Prior Knowledge</i>	<i>Actions</i>	<i>Percept Sequence</i>
ضابطه‌ی موفقیت عامل	دانایی پیشینی عامل از محیط	کنش‌هایی که عامل قادر به انجام آنهاست.	دنباله‌ی ادراکی عامل تا آن زمان



عامل رسیونال

RATIONAL AGENT

یک عامل رسیونال

کنشی را انتخاب می‌کند که
مقدار مورد انتظار معیار کارآیی را
با داشتن دنباله‌ی ادراکی تا آن لحظه
ماکزیم می‌کند.

عامل رسیونال

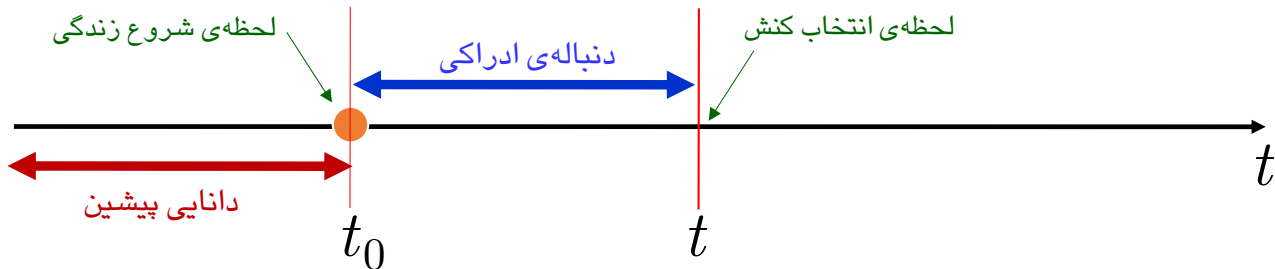
تعریف (راسل و نوریگ)

RATIONAL AGENT

یک عامل رسیونال

برای هر دنباله‌ی ادراکی ممکن
کنشی را انتخاب می‌کند که
انتظار دارد معیار کارآیی آن را ماکزیمم کند؛

بر اساس شواهدی که توسط دنباله‌ی ادراکی فراهم می‌شود و
آنچه دانایی درون‌سازی شده‌ی عامل است.



رسیونالیته

RATIONALITY

پیامد کنش‌ها ممکن
است مورد انتظار
نباشد

آگاه از غیب

Clairvoyant

ادراکات ممکن است
همه‌ی اطلاعات
مربوط را فراهم نکند

همه‌چیزدان

Omniscient

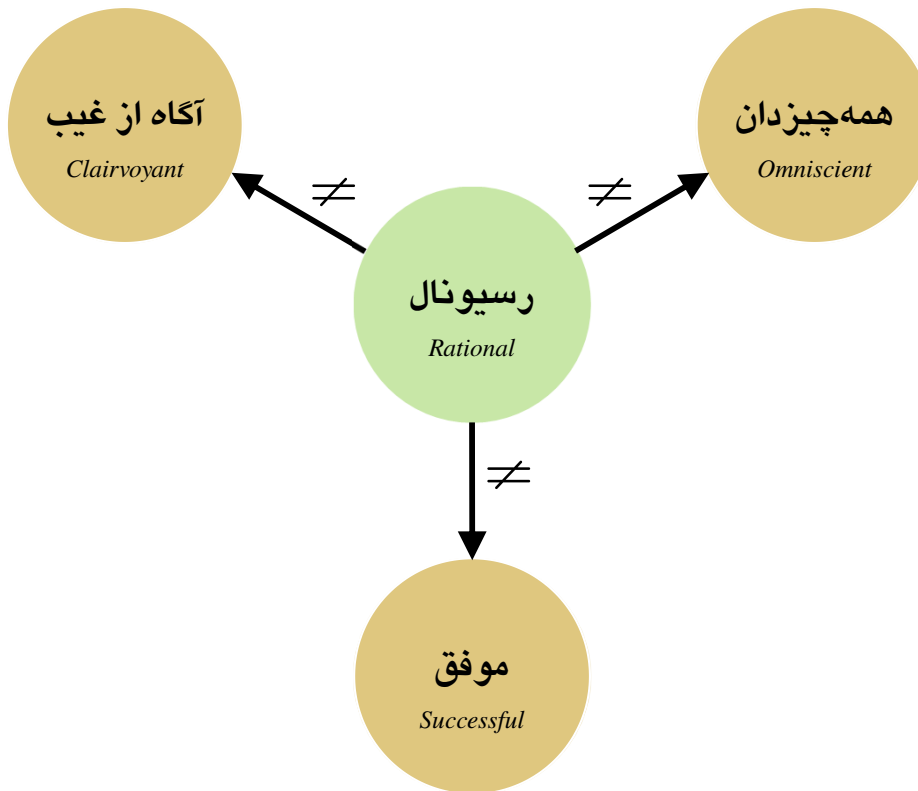
رسیونال

Rational

موفق

Successful

عامل ممکن است
شکست بخورد



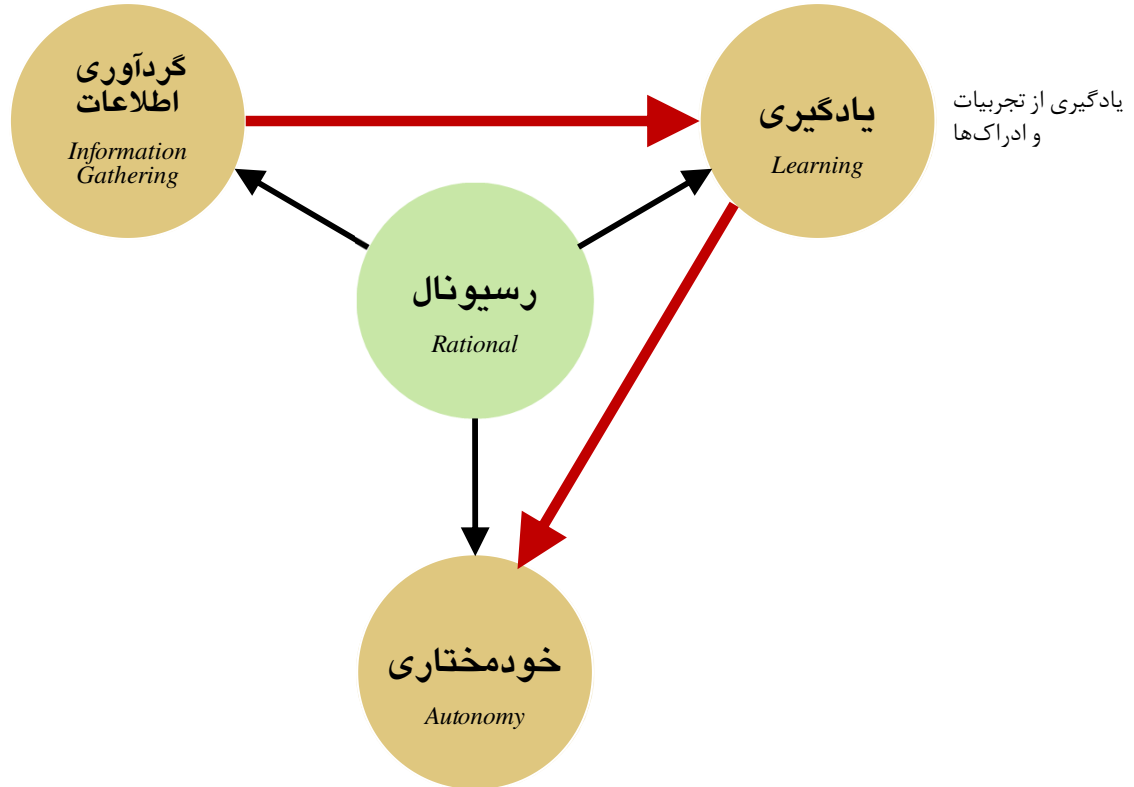
تمایز رسیونالیته با کامل بودن

کامل بودن <i>Perfection</i>	رسیونالیته <i>Rationality</i>
ماکزیمم‌سازی کارآیی واقعی (actual)	ماکزیمم‌سازی کارآیی مورد انتظار (expected)
انجام بهترین کار ممکن	انجام بهترین کاری که توانسته است بفهمد
پیاپاده‌سازی غیر ممکن	پیاپاده‌سازی امکان‌پذیر

رسیونالیته

ملزومات

عامل نباید رفتار غیر
هوشمندانه داشته باشد:
قبل از تصمیم‌گیری باید
اطلاعات کافی از محیط
جمع کند



یادگیری از تجربیات
و ادراک‌ها

دانایی عامل مستقل از دانایی اولیه‌ی آن می‌شود؛
رفتار عامل توسط تجربه‌ی او تعیین می‌شود.

هوش مصنوعی و عامل های هوشمند

۴

محیط

مشخص‌سازی محیط وظیفه

SPECIFYING THE TASK ENVIRONMENT

P	E	A	S
معیار کارآیی	محیط	کنش‌گرها	حسگرها
<i>Performace Measure</i>	<i>Environment</i>	<i>Actuators</i>	<i>Sensors</i>

مشخص‌سازی محیط وظیفه

مثال: دنیای جاروبرقی

SPECIFYING THE TASK ENVIRONMENT: THE VACUUM-CLEANER AGENT

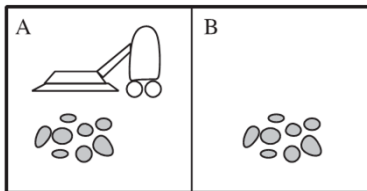
P	E	A	S
معیار کارآیی	محیط	کنش‌گرها	حسگرها
<i>Perfomance Measure</i>	<i>Environment</i>	<i>Actuators</i>	<i>Sensors</i>

برای هر مربع تمیز
در هر گام زمانی، یک
امتیاز مثبت
کلاً ۱۰۰۰ گام زمانی
طول عمر

نقشه‌ی محیط معلوم،
توزیع آشغال‌ها و
مکان اولیه‌ی عامل
نامعلوم

انجام کنش‌های
راست
چپ
مکش
هیچ

حسگر مکان
حسگر کثیفی مربع



مشخص‌سازی محیط وظیفه

مثال: تاکسی خودکار هوشمند

SPECIFYING THE TASK ENVIRONMENT: AUTOMATIC INTELLIGENT TAXI

P	E	A	S
معیار کارآیی <i>Performace Measure</i>	محیط <i>Environment</i>	کنش‌گرها <i>Actuators</i>	حسگرها <i>Sensors</i>
ایمنی رسیدن به مقصد فایده رعایت قانون راحتی ...	خیابان‌ها/آزادراه‌ها ترافیک علایم راهنمایی عابرین پیاده آب و هوا ...	فرمان گاز ترمز بوق بلندگو/نمایشگر ...	تصویر شتاب‌سنج درجه‌ها حسگرهای موتور صفحه‌کلید موقعیت‌سنج GPS ...

مشخص‌سازی محیط وظیفه

مثال: عامل خرید اینترنتی

SPECIFYING THE TASK ENVIRONMENT: INTERNET SHOPPING AGENT

P	E	A	S
معیار کارآیی <i>Performace Measure</i>	محیط <i>Environment</i>	کنش‌گرها <i>Actuators</i>	حسگرها <i>Sensors</i>
قیمت کیفیت مناسب بودن کارآمدی ...	سایت‌های وب حال و آینده فروشندگان خریداران ...	نمایش به کاربر دنبال کردن یک URL پر کردن فرم ...	خواننده‌ی صفحات HTML (متن، گرافیک، اسکریپت) ...

خصوصیات محیط‌های وظیفه

PROPERTIES OF TASK ENVIRONMENTS

مشاهده‌پذیر کامل <i>Fully Observable</i>	مشاهده‌پذیر جزئی <i>Partially Observable</i>
تک عاملی <i>Single-agent</i>	چند عاملی <i>Multiagent</i>
قطعی <i>Deterministic</i>	استراتژیک <i>Strategic</i>
مقطعی <i>Episodic</i>	اتفاقی <i>Stochastic</i>
ایستا <i>Static</i>	دنباله‌ای <i>Sequential</i>
گسسته <i>Discrete</i>	پویا <i>Dynamic</i>
شناخته‌شده <i>Known</i>	پیوسته <i>Continuous</i>
	ناشناخته <i>Unknown</i>

خصوصیات محیط‌های وظیفه

مشاهده‌پذیر کامل یا مشاهده‌پذیر جزئی

مشاهده‌پذیر کامل *Fully Observable*

حسگرهای عامل
دسترسی به حالت
کامل محیط در هر
لحظه را به عامل
می‌دهند.

راحتی کار: عامل برای
دنبال کردن دنیا نیازی
به نگهداری حالت
داخلی ندارد.

مشاهده‌پذیر کامل به طور مؤثر *effectively fully observable*

حسگرهای عامل

همه‌ی جنبه‌های مربوط
به انتخاب کنش را
نشان می‌دهند.

(مربوط بودن وابسته به معیار کارایی)

مشاهده‌پذیر جزئی *Partially Observable*

حسگرهای عامل
دسترسی به جزئی
از حالت محیط در هر
لحظه را به عامل
می‌دهند.

مثلا عدم وجود حسگر،
حسگرهای نادقیق،
حسگرهای نویزی

مشاهده‌ناپذیر *unobservable*

عامل هیچ حسگری ندارد.

خصوصیات محیط‌های وظیفه

تک عاملی یا چند عاملی

تک عاملی

Single-agent

یک عامل به تنهایی در محیط عمل می‌کند.

چند عاملی

Multiagent

چند عامل در محیط عمل می‌کنند:

همکارانه

Cooperative

افزایش معیار کارایی
یک عامل باعث
افزایش معیار کارایی
عامل دیگر می‌شود.

مثال: گروه سرود

نیمه ...

Partially ...

مثال: محیط رانندگی:
جای پارک: رقابتی
اجتناب از تصادف: همکارانه

رقابتی

Competitive

افزایش معیار کارایی
یک عامل باعث کاهش
معیار کارایی عامل
دیگر می‌شود.

مثال: بازی شطرنج

استفاده از دانایی عامل‌های دیگر

برقراری ارتباط
communication

اجتناب از مضرات پیش‌بینی‌پذیری

رفتار تصادفی‌شده
randomized behavior

رفتارهای رسیونال
خاص محیط‌های چندعاملی

کدام موجودیت می‌تواند به عنوان عامل دیده شود؟ ← هر چیزی که محیط را درک کند و روی محیط کنش انجام دهد.

کدام موجودیت می‌باید به عنوان عامل دیده شود؟ ← هر چیزی که برای ماکزیم‌سازی معیار کارایی‌اش، که به رفتار دیگری هم وابسته است، تلاش می‌کند.

خصوصیات محیط‌های وظیفه

قطعی یا اتفاقی

قطعی

Deterministic

حالت بعدی محیط توسط
حالت فعلی و کنش اجرا
شده توسط عامل به طور
کامل تعیین می‌شود.

راحتی کار: عامل برای دنبال
کردن دنیا نیازی به نگهداری
حالت داخلی ندارد.

استراتژیک

Strategic

محیط قطعی است، بجز در
مورد کنش سایر عامل‌ها

اتفاقی

Stochastic

حالت بعدی محیط توسط
حالت فعلی و کنش اجرا
شده توسط عامل به طور
کامل تعیین نمی‌شود.

- محیط مشاهده‌پذیر جزئی
ممکن است اتفاقی به نظر برسد.
- عدم قطعیت در مورد برآمدها
برحسب احتمالات کمی می‌شوند.

نامطمئن

Uncertain

محیطی که مشاهده‌پذیر
کامل یا قطعی نباشد.

غیر قطعی

Non-deterministic

حالت بعدی محیط توسط حالت فعلی و کنش
اجرا شده توسط عامل به طور کامل تعیین نمی‌شود.

کنش‌ها با برآمدهای ممکن آنها مشخص می‌شوند،
بدون انتساب احتمال به آنها

خصوصیات محیط‌های وظیفه

مقطعی یا دنباله‌ای

مقطعی

Episodic

تجربه‌ی عامل قابل تقسیم به مقطع‌های اتمیک است.

در هر مقطع، عامل یک ادراک دریافت می‌کند و سپس یک کنش واحد انجام می‌دهد.

مقطع بعدی به کنش‌های انجام شده در مقاطع قبلی وابسته نیست.

انتخاب کنش در هر مقطع فقط وابسته به همان مقطع است.

ساده‌تر است: عامل نیازی ندارد به جلو فکر کند!

مثال: عمده‌ی وظایف طبقه‌بندی

دنباله‌ای

Sequential

تصمیم فعلی می‌تواند بر همه‌ی تصمیم‌های آینده تأثیر بگذارد.

کنش‌های کوتاه‌مدت می‌توانند پی‌آمدهای بلندمدت داشته باشند.

مثال: شطرنج، رانندگی تاکسی

خصوصیات محیط‌های وظیفه

ایستا یا پویا

ایستا

Static

● اگر محیط نتواند در هنگام تأمل عامل تغییر کند.

محیط برای آن عامل، ایستا است.

ساده‌تر است: (۱) عامل نیازی ندارد در هنگام تصمیم‌گیری در مورد یک کنش به نگاه کردن ادامه دهد. (۲) عامل لازم نیست نگران گذر زمان باشد.

مثال: شطرنج معمولی

نیمه‌پویا

Semidynamic

● اگر خود محیط با گذر زمان تغییر نکند، اما امتیاز کارآیی عامل تغییر کند.

مثال: شطرنج با ساعت

پویا

Dynamic

● اگر محیط بتواند در هنگام تأمل عامل تغییر کند.

محیط برای آن عامل، پویا است.

محیط پویا به طور مداوم از عامل می‌پرسد که می‌خواهد چه کنشی را انجام دهد. اگر هنوز تصمیم نگرفته باشد، فرض می‌کند تصمیم گرفته است کاری انجام ندهد.

مثال: رانندگی تاکسی (سایر خودروها و عابرین در حین تصمیم‌گیری عامل، حرکت می‌کنند).

خصوصیات محیط‌های وظیفه

گسسته یا پیوسته

گسسته

Discrete

مؤلفه‌های گسسته
در کار هستند.

مثال: شطرنج معمولی
(حالت، زمان، ادراک،
کنش)

پیوسته

Continuous

مؤلفه‌های پیوسته
در کار هستند.

مثال: رانندگی تاکسی
(حالت، زمان، ادراک،
کنش)

تمایز پیوسته و گسسته: در مؤلفه‌های
حالت محیط، نحوه‌ی برخورد با زمان، ادراک‌ها و کنش‌های عامل

خصوصیات محیط‌های وظیفه

شناخته‌شده یا ناشناخته

شناخته‌شده

Known

برآمدها (یا احتمال برآمدها
در محیط اتفاقی) برای هم‌ی
کنش‌ها داده شده است.

یک محیط شناخته‌شده می‌تواند
مشاهده‌پذیر جزئی باشد.

مثال: بازی کارت

تمایز میان محیط شناخته‌شده
و ناشناخته، همان تمایز بین
محیط مشاهده‌پذیر کامل و
جزئی نیست.

ناشناخته

Unknown

عامل باید یاد بگیرد که چگونه
عمل کند تا تصمیم‌های خوبی
بگیرد.

یک محیط ناشناخته می‌تواند
مشاهده‌پذیر کامل باشد.

مثال: بازی‌های ویدئویی جدید؛
صفحه‌ی بازی می‌تواند کل حالت بازی را نشان
بدهد، اما معلوم نیست دکمه‌ها چه کاری انجام
می‌دهند که با آزمایش معلوم می‌شود.

این خصوصیت، فقط به خود محیط بر نمی‌گردد، بلکه
به حالت دانایی عامل (یا طراح آن) در مورد «قوانین فیزیک» محیط بر می‌گردد.

خصوصیات محیط‌های وظیفه

مثال‌هایی از محیط‌های وظیفه و مشخصه‌های آنها

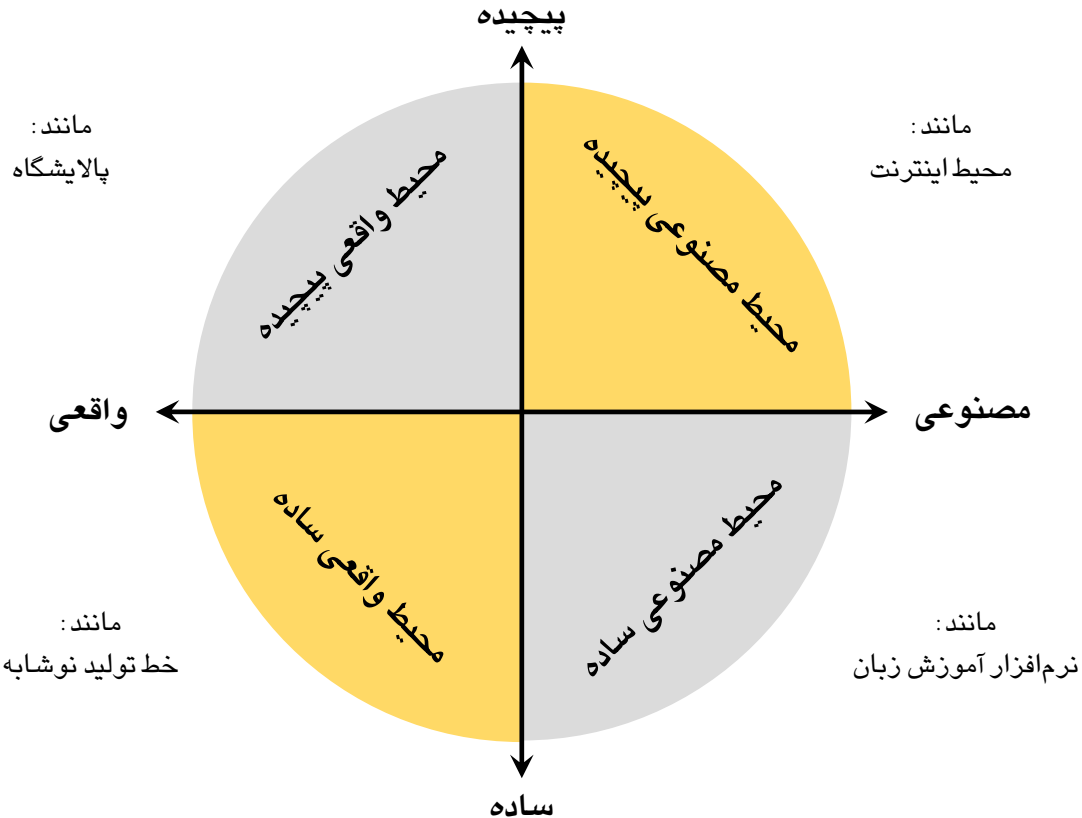
مشاهده‌پذیر؟	چندعاملی؟	قطعی؟	مقطعی؟	ایستا؟	گسسته؟
کامل	چندعاملی	قطعی	دنباله‌ای	ایستا	گسسته
کامل	چندعاملی	اتفاقی	دنباله‌ای	ایستا	گسسته
جزئی	تک‌عاملی*	تاحدودی	دنباله‌ای	نیمه‌پویا	گسسته
جزئی	چندعاملی	اتفاقی	دنباله‌ای	پویا	پیوسته
جزئی	تک‌عاملی	اتفاقی	دنباله‌ای	پویا	پیوسته
جزئی	تک‌عاملی	اتفاقی	دنباله‌ای	پویا	پیوسته

نوع محیط، روش طراحی عامل را مشخص می‌کند.

هر مجموعه از روش‌های هوش مصنوعی، برای طراحی عامل در نوع خاصی از محیط مناسب است.

انواع محیط

نسبت محیط ساده / پیچیده با محیط واقعی / مصنوعی

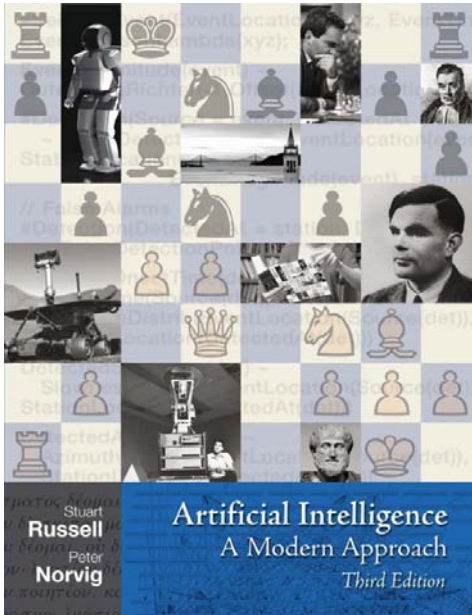


هوش مصنوعی و عوامل های هوشمند

۵

منابع

منبع اصلی



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

Chapters 1, 2

1 INTRODUCTION

In which we try to explain why we consider artificial intelligence to be a subject most worthy of study, and in which we try to decide what exactly it is, this being a good thing to decide before embarking.

INTELLIGENCE

We call ourselves *Homo sapiens*—man the wise—because our **intelligence** is so important to us. For thousands of years, we have tried to understand *how we think*; that is, how a mere handful of matter can perceive, understand, predict, and manipulate a world far larger and more complicated than itself. The field of **artificial intelligence**, or AI, goes further still: it attempts not just to understand but also to *build* intelligent entities.

ARTIFICIAL INTELLIGENCE

AI is one of the newest fields in science and engineering. Work started in earnest soon after World War II, and the name itself was coined in 1956. Along with molecular biology, AI is regularly cited as the “field I would most like to be in” by scientists in other disciplines. A student in physics might reasonably feel that all the good ideas have already been taken by Galileo, Newton, Einstein, and the rest. AI, on the other hand, still has openings for several full-time Einsteins and Edisons.

AI currently encompasses a huge variety of subfields, ranging from the general (learning and perception) to the specific, such as playing chess, proving mathematical theorems, writing poetry, driving a car on a crowded street, and diagnosing diseases. AI is relevant to any intellectual task; it is truly a universal field.

1.1 WHAT IS AI?

RATIONALITY

We have claimed that AI is exciting, but we have not said what it *is*. In Figure 1.1 we see eight definitions of AI, laid out along two dimensions. The definitions on top are concerned with *thought processes and reasoning*, whereas the ones on the bottom address *behavior*. The definitions on the left measure success in terms of fidelity to *human* performance, whereas the ones on the right measure against an *ideal* performance measure, called **rationality**. A system is rational if it does the “right thing,” given what it knows.

Historically, all four approaches to AI have been followed, each by different people with different methods. A human-centered approach must be in part an empirical science, in-