

CCAI 2020 大会特邀报告(II)

中国科大陈小平教授：

图灵假说70年：两类AI与封闭性挑战

临菲信息技术港

2020年8月30日，CCAI 2020 第二天大会特邀报告。ppt 由本号根据视频摘录。

报告人：陈小平，中国科学技术大学机器人实验室主任、教授



一、图灵假说与两类人工智能

Turing's Hypothesis and Two Sorts of AI

二、两种AI经典思维与封闭性

Two Thinking Modes of AI and the Closeness Properties

三、封闭性的基础研究挑战

The Closeness Challenge to Basic Research in AI

四、超越封闭性：开放知识技术路线

The Open-knowledge Approach to exceeding the Closeness

一、图灵假说与两类人工智能

Turing's Hypothesis and Two Sorts of AI

Godel-Turing Corollary in Prehistoric Basic Research of AI:
Complex Reasoning Can Be Implemented as Turing Computation

陈小平 (中国科学技术大学)

人工智能史前基础研究回顾

- ❖ 约300B.C., Euclid《几何原本》: 第一个**实质公理系统**。
- ❖ 1899年Hilbert《几何基础》: 第一个**形式公理系统**。现代逻辑诞生。
- ❖ 1651年Hobbes《利维坦》: 推理与计算可以相互转化。
- ❖ 1931年Godel不完备性定理的中间结果: K_N 可表示和**递归函数**等价。
- ❖ 1937年Turing提出图灵机模型和Church-Turing论题——可计算的都是图灵机可计算的。计算机科学的诞生。
- ❖ **Godel-Turing推论[被严格证明] 大量复杂推理(如命题演算可表达的推理、 K_N 可表示推理)是图灵可计算的, 即可转化为图灵机上的计算。**

陈小平(主编), 《人工智能伦理导引》, 第2章, 中国科学技术大学出版社, 2020年(待出)。

中国人工智能大会, 南京, 2020年8月29日

Turing's Hypothesis: Decision, Learning, Understanding,
Creation, etc. Can Be Implemented as Turing Computation

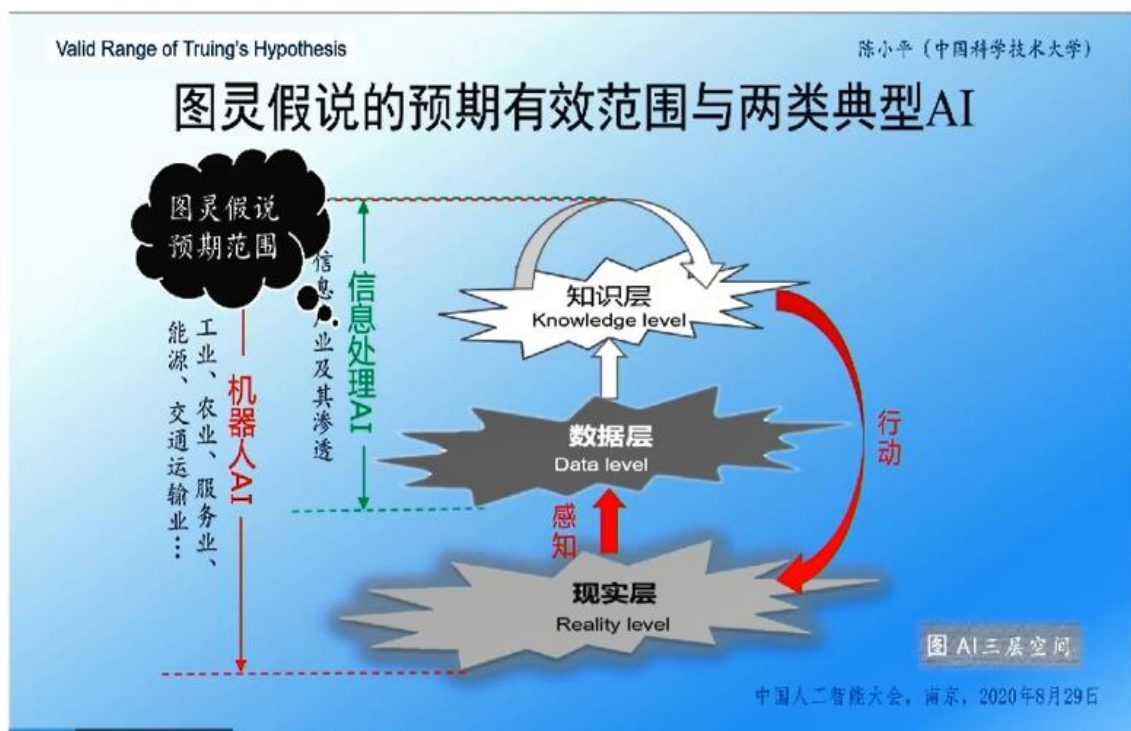
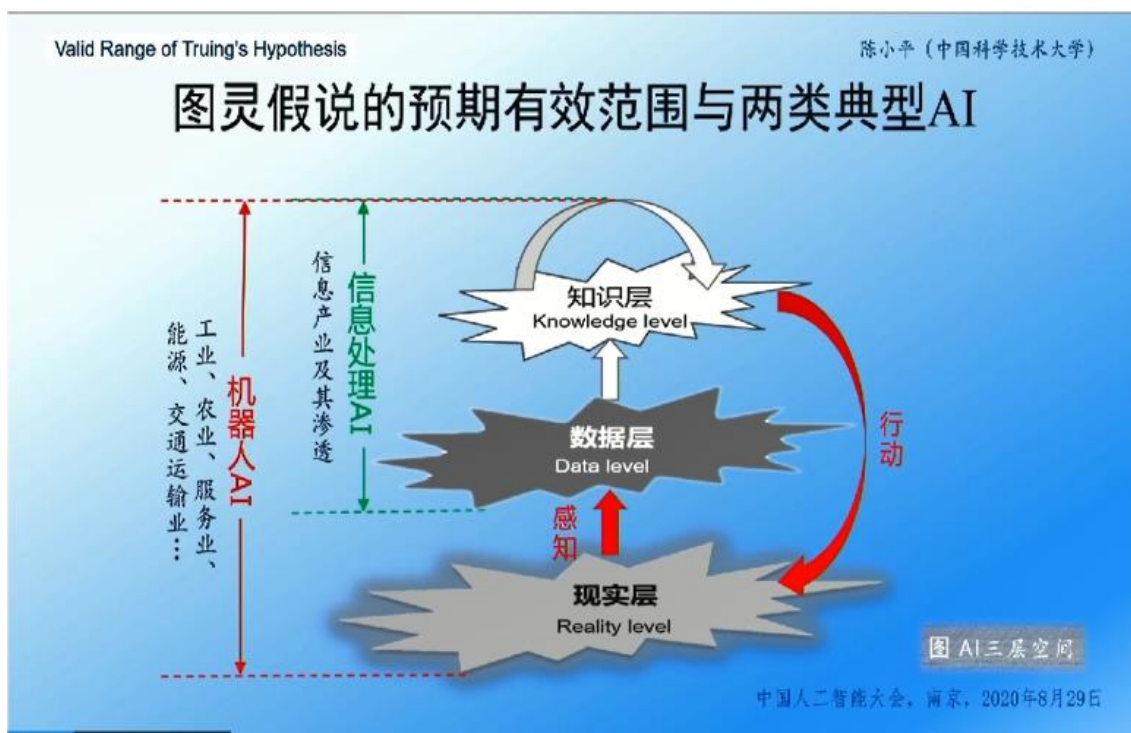
陈小平 (中国科学技术大学)

图灵假说

- ❖ **图灵测试** 1950年《Mind》论文提出图灵测试, 至今**70周年**。
- ❖ 科学试验的目的通常是证实/证伪科学假说。图灵测试的目的?
- ❖ **图灵假说** 不仅推理, 而且**决策、学习、理解、创造等人类智力活动都可归结为图灵计算**。
- ❖ 图灵假说是Godel-Turing推论的实质性推广, 人类科技史上最伟大的假说之一, 而图灵测试是其验证手段。
- ❖ **现实意义** 当时第一批电子数字计算机已研制成功并投入使用。如果图灵假说成立, 上述种类的智能都可以在计算机上实现。
- ❖ **图灵假说的预期有效范围?**

A. M. Turing, Computing Machinery and Intelligence. Mind 49: 433-460, 1950.

中国人工智能大会, 南京, 2020年8月29日



Two Sorts of AI: Examples

陈小平 (中国科学技术大学)

两类AI: 例子

信息处理AI	机器人AI
❖ 例1 聊天机器人	❖ 例2 家庭服务机器人
用户: 请帮我端饭。	用户: 请帮我端饭。
机器人: 自己端。	机器人: 自己端。
用户: 有智能!	用户: 有个性!
用户: 请帮我端汤。	用户: 请帮我端汤。
机器人: 自己端。	机器人: 自己端。
用户: 你会的句型有点少。	用户: 你只会聊天。退货!

陈小平(主编), 《人工智能伦理导引》, 第2章, 中国科学技术大学出版社, 2020年(待出).

FAIR' 2020: Grand Challenges and Opportunities, 19/7/2020

Turing's Plan for the Two Sorts of AI

陈小平 (中国科学技术大学)

图灵关于两类AI的观点

- 1948年, 在亲自打印的手稿Intelligent Machinery中, 图灵首先探讨了如何建造用机器替代人的每一部分intelligent machinery, 如无人车。
- 考虑到当时的技术条件, 图灵建议先研究没有感知和行动能力的thinking machines——信息处理AI。
- 1950年提出的“图灵测试”只针对信息处理AI, 并建议研究围棋等博弈、语言学习、机器翻译等。
- 图灵实际上考虑了两类人工智能: 第一阶段ITAI, 第二阶段RAI。



Intelligent Machinery手稿第20~21页

中国人工智能大会, 南京, 2020年8月29日

二、两种AI经典思维与封闭性

Two Thinking Modes of AI and the Closeness Properties

The First Thinking Mode of AI: Brute-force

陈小平 (中国科学技术大学)

第一种AI经典思维：强力法

❖ 强力法主要成分：

1. 模型(知识表示、搜索空间...)
2. 推理机或搜索算法
3. 优化(模型优化、搜索算法优化...)

❖ 基础性数学工具：逻辑、概率、决策论规划

❖ 搜索算法：蒙特卡洛树搜索、A*算法、.....

❖ 机器人AI的模型构成[信息处理AI通常只有KB]：

知识库KB(抽象知识)+模型降射 g^m (KB到现实场景的grounding)

哲学与认知科学明德讲坛第11期“AI与情感”，人大，2020年7月17日

Big Knowledge: The Trend in the Brute-force Thinking since 1990s 陈小平 (中国科学技术大学)

大知识——1990s以来强力法的新赛道

◆ 大型知识基础工程：大型**知识体**建造、抽取和推理。代表性工作：



Cyc (Lenat, 1984-): A comprehensive ontology and knowledge base. CycL (higher order logic). 416,000 collections (types, sorts, etc), 42,500 predicates (relations, attributes, functions, etc).

Open Mind Common Sense (MIT, 1999-): A project to build and utilize a large **commonsense knowledge base** from Web users. It is built on the **structured natural language corpus**, a semantic network, ConceptNet, as well as a matrix-based representation, AnalogySpace.

Semantic Web (Tim Berners-Lee, 2001-): A set of **standards** for common data formats and exchange protocols on the Web, which allow data to be shared and reused across application, enterprise, etc.



Freebase (2007-16): A collaborative knowledge base from many sources with **3B facts and costs \$6.75B**.

IBM Watson (2011-): The sources of information for Watson include encyclopedias, dictionaries, thesauri, newswire articles and literary works, databases, taxonomies and ontologies.

Knowledge Graph (2010-): A large knowledge base organized as a graph. **FrameNet, WordNet, Ontologies, ...**

杰出校友数学论坛，中科大数学学院，2019年3月6日

Model, Reasoning and Grounding in AI 陈小平 (中国科学技术大学)

强力法示例：模型、推理与降射

◆ 例3 就餐服务的AI模型

1. 知识库(一阶逻辑公式表达知识)	直观解释
$\forall x, y (dish(x) \wedge food(y) \rightarrow hold(x, y))$	餐具可以盛食物
$food(rice)$	米饭是食物
$food(soup)$	汤是食物
$dish(bowl)$	碗是餐具

2. 模型降射 (Modeled Grounding)



推理机(程序)

提问: $hold(bowl, x)?$
 $hold(bowl, soup)?$
 $hold(bowl, rice)?$

回答: $rice, soup...$
 Yes
 Yes

降射是RAI核心挑战

知识库+模型降射

FAIR' 2020: Grand Challenges and Opportunities, 19/7/2020

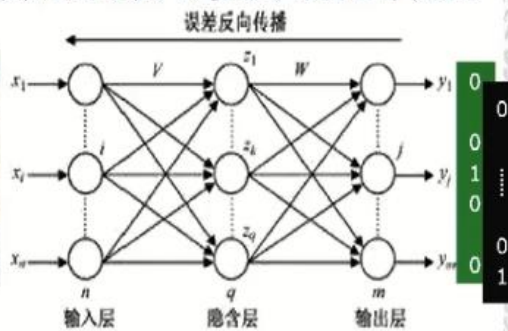
The Second Thinking Mode of AI: Training

陈小平 (中国科学技术大学)

第二种AI经典思维：训练法

- ❖ 计算过程：对每一个“输入”用人工神经网络计算对应的“输出”。
- ❖ 训练过程：采集样本，用监督学习算法训练神经网络，尽量缩小输出整体误差。
- ❖ 例4 ImageNet比赛图像分类任务：

数 据	人 工 标 注
7种鱼的图片	0 - 6
公鸡、母鸡图片	7 - 8
26种鸟的图片	9 - 34
.....	
卫生纸图片	999



- ❖ 降维 采集所有代表性数据并人工标注，训练网络以建立完全的模型降维和行动策略。
- ❖ 挑战 哪些是代表性数据？

ImageNet a large visual database for use in visual object recognition. More than 14 million images have been hand-annotated.
<https://en.wikipedia.org/wiki/ImageNet>

第15届中国软科学学术年会，北京，2019年11月23日

Three Long-standing Puzzles Stemming from Grounding in AI Research since 1980s

陈小平 (中国科学技术大学)

1980年代以来的三个难点

非典型情况穷尽了吗?

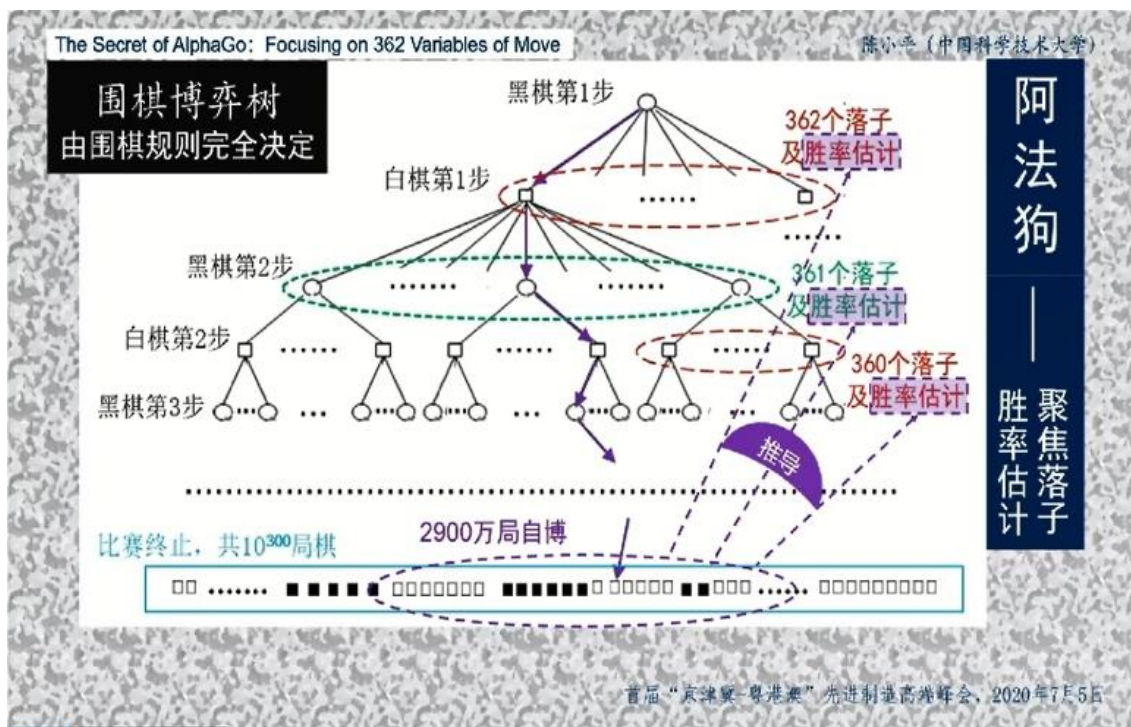
非典型情况都可解吗?

非典型情况会致命吗?

The Brittleness Problem When an intelligent system is confronted by unanticipated situations which do not fit the internal model of the system, the system is likely to reach wrong conclusions or even breaks.

图片来自网络

中国人工智能大会，南京，2020年8月29日



AlphaGo: Integration of Brute Force and Training

陈小平 (中国科学技术大学)

AlphaGo的技术构成

AlphaGo三代战胜所有人类围棋高手, 四代100:0三代
人工智能应该模拟人的思维?

AlphaGo Zero的四项核心技术

1. 简化的决策论规划模型
2. 蒙特卡洛树搜索
3. 深层神经网络 (残差网络)
4. 强化学习

强力法(推理+搜索)

训练法

Mastering the game of Go without human knowledge, *Nature*, Oct. 18, 2017

清华大学, 2019年9月27日

The Closeness Principles for Effectiveness of the Two Modes

陈小平 (中国科学技术大学)

封闭性准则：强力法、训练法的有效条件

	对应用场景的要求	实例：AlphaGo Zero
强力法	用 确定 的一组变元， 完全描述 应用场景	变元的确定集合：362个落子
	变元服从 领域定律 并可用 AI模型 表达	围棋规则+自博⇒落子的胜率估计
	AI模型的预测与应用场景足够接近	实战全胜
训练法	存在完整、确定的设计和评价准则	战胜所有对手
	存在足够好的 代表性数据集	2900万局自博数据及自动标注
	神经网络经过训练符合评价准则	实战全胜

陈小平，人工智能中的封闭性和强封闭性——现有成果的能力边界、应用条件和伦理风险，《智能系统学报》2020年第1期/CAAI Trans. Intelligent Systems, 15(1): 114-120.

哲学与认知科学明德讲坛第11期“AI与儒学”，人大，2020年7月17日

Applicability of the Two Modes in Current Industries

陈小平 (中国科学技术大学)

强力法和训练法的应用条件

强力法和训练法
在封闭性或封闭化场景中的应用不存在理论困难

符合条件的**现有应用场景**：
制造业、智慧农业、物流业、IT业、部分服务业、……
产业落地路径：
封闭化、半封闭化、柔性化、……

陈小平，封闭性场景：人工智能的产业化路径，《文化纵横》2020年第1期

北京大学，2019年9月25日

Application Path 1: Close a Scenario Completely

陈小平 (中国科学技术大学)

落地路径1：应用场景完全封闭化



自然生产过程
是非封闭性
/非结构化的

改造为封闭性
/结构化的自动
化生产过程

参考案例：汽车生产线

图片来源自网络

第五届中国服务机器人产业高峰论坛，南京，2019年6月29日

Application Path 2: Close a Compound Scenario Partially

陈小平 (中国科学技术大学)

落地路径2：应用场景半封闭化



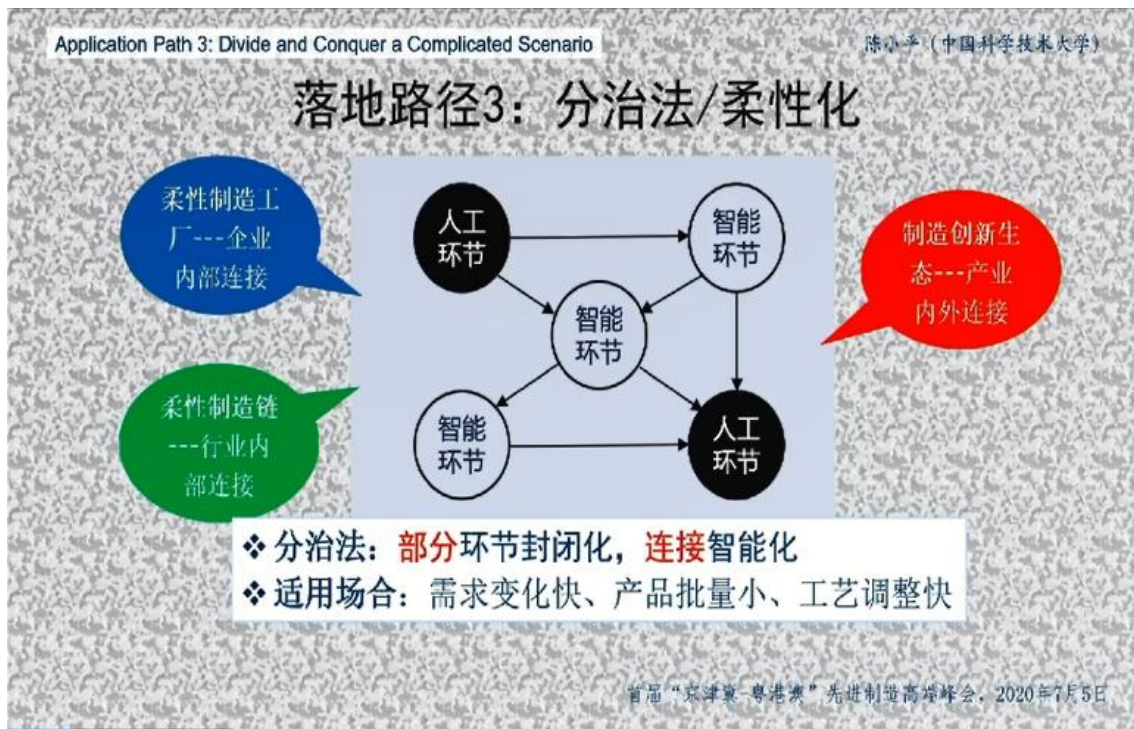

高铁列车行车
过程全封闭化

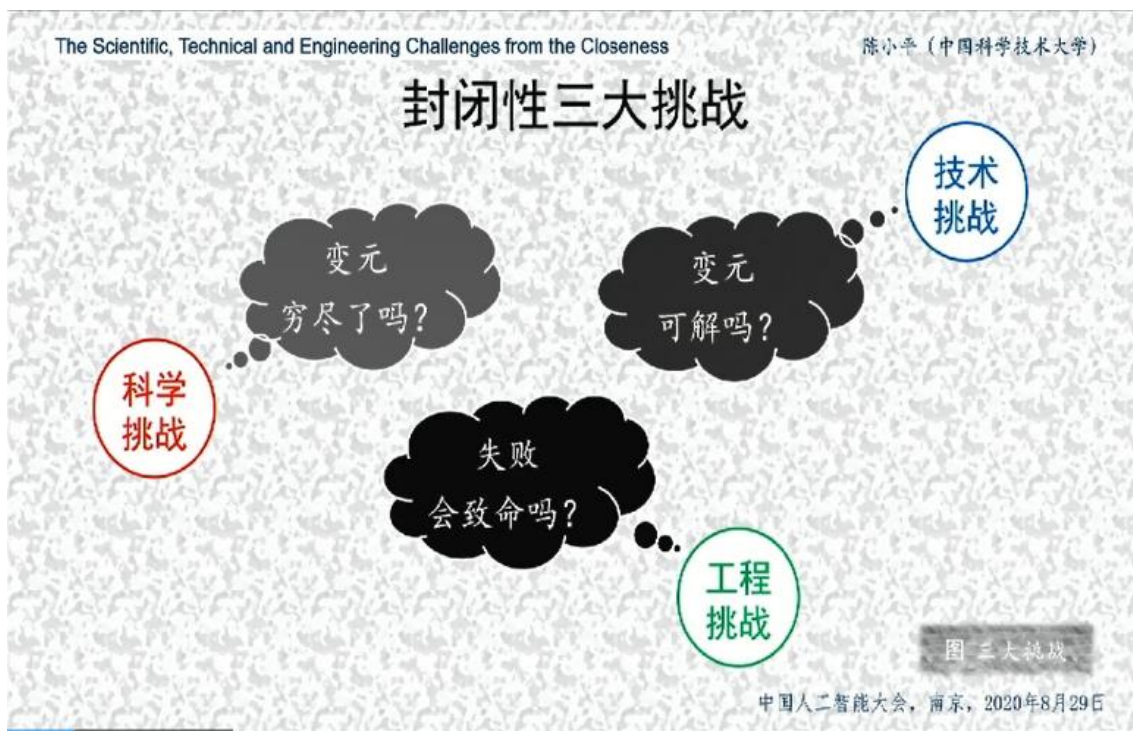
站台、候车室、
车厢非封闭化

参考案例：高铁

图片来源自网络

北京大学，2019年9月25日





Scientific Challenge: Endlessness of Grounding

陈小平 (中国科学技术大学)

封闭性的科学挑战：无尽性

- ❖ 变元 [逻辑] 个体符号指称场景对象，谓词符号指称对象属性，对象和属性通称场景元素。知识库 KB 所包含的每一个个体符号和每一个谓词符号称为 KB 的一个变元。
- ❖ 降射 g 任给场景 S 和变元 x ， $g(x)$ 是变元 x 在 S 中对应的所有场景元素的集合。
- ❖ 模型降射 g^m 降射 g 被 AI 模型建模的部分。
- ❖ 丢失变元 $g^m(x) \neq g(x)$ 的变元 x ；称不完全 g^m 。

例3(续)

$g^m(\text{bowl})$:

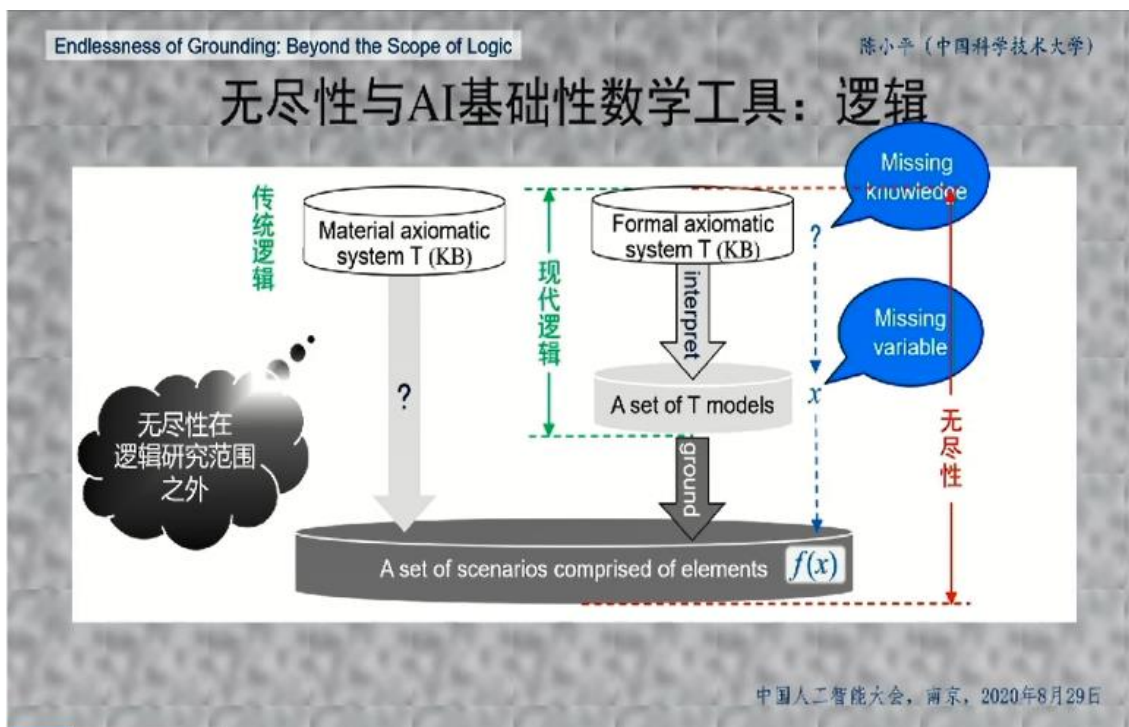
$g(\text{broken bowl})$:

不完全 g^m

丢失变元

- ❖ 无尽性 如果没有判断一个场景 S 在一个 AI 模型 M 下是否存在丢失变元的系统性可行方法，则称 S 的降射 g 是无尽的，称 S 在 M 下是非封闭的。

中国人工智能大会，南京，2020年8月29日



Endlessness of Grounding: Beyond the Scope of Probability / Decision-theoretic Planning

陈小平 (中国科学技术大学)

无尽性与AI基础性数学工具：概率/决策论规划

- ❖ 在概率形式化下，每一个样本点 $\omega \in \Omega$ 是一个变元， $f(\omega)$ 是 ω 对应的场景元素。无尽性意味着 Ω 中变元可能未穷尽所有场景元素。
- ❖ Kolmogoroff 公理 (1933)
 1. $P(A) \geq 0$, where A is any event in Ω ;
 2. $P(\Omega) = 1$, where Ω is the sample space;
 3. $P(A+B) = P(A) + P(B)$, where A and B are mutually exclusive.
- ❖ 分析 Kolmogoroff 概率论只考虑给定样本空间 Ω 中的概率推理；增加新变元 $\omega' \notin \Omega$ 可能导致对所有 $\omega \in \Omega$, $P(\omega)$ 的值发生改变。
- ❖ 观察 无尽性在Kolmogoroff 概率论的研究范围之外。

中国人工智能大会，南京，2020年8月29日

The Technical Challenge: Hard-to-solve Variables

陈小华 (中国科学技术大学)

封闭性的技术挑战：难解性

深层网络
识别结果

❖ 难解变元 现实世界中的真实场景往往存在着各种难以识别、难以求解的变元，如视觉识别中的遮挡、光照变化等。

❖ 例5 “遮挡”是一个**无限值变元**，每个值**降射**为一种可能的相对位置组合。



Alan L. Yuille & Chenxi Liu, Limitations of Deep Learning for Vision, and How We Might Fix Them, *The Gradient*, 2018.

中国人工智能大会，南京，2020年8月29日

Hard-to-solve Variables Also Studied in Statistics

陈小华 (中国科学技术大学)

统计学中的难解变元：Latent Variables

❖ 8种类型均在AI中出现，通常不存在通用解法：

- ‘True’ variables measured with error
- Hypothetical constructs
- Unobserved heterogeneity
- Missing data
- Counterfactuals or ‘potential outcomes’
- Latent responses underlying categorical variables
- Generate flexible multivariate distributions
- Combine information about individual units from different sources

Ad hoc solutions
based on detailed
analysis of a
concrete scenario

Anders Skrondal and Sophia Rabe-Hesketh. Generalized latent variable modeling, Chapman & Hall/CRC, 2004.

中国人工智能大会，南京，2020年8月29日

The Engineering Challenge: Fatality and Solutions to It

陈小平 (中国科学技术大学)

封闭性的工程挑战：失误致命性

- ❖ **失误致命性** 如果在一个场景S中，丢失变元和难解变元的存在导致AI系统出现**不可接受的失误**，则称S是失误致命的。
- ❖ **致命性的现有应对策略** 封闭化、半封闭化、柔性化(第二节)。
- ❖ **封闭化** 对一个失误致命的场景S进行场景裁剪和场景改造，使得在裁剪或改造后的场景S'中，不存在任何丢失变元或没有得到有效处理的难解变元的过程，称为**场景封闭化**。

陈小平，封闭性场景：人工智能的产业化路径，《文化纵横》2020年第1期

中国人工智能大会，南京，2020年8月29日

Definition of Closeness and Its Significance in AI

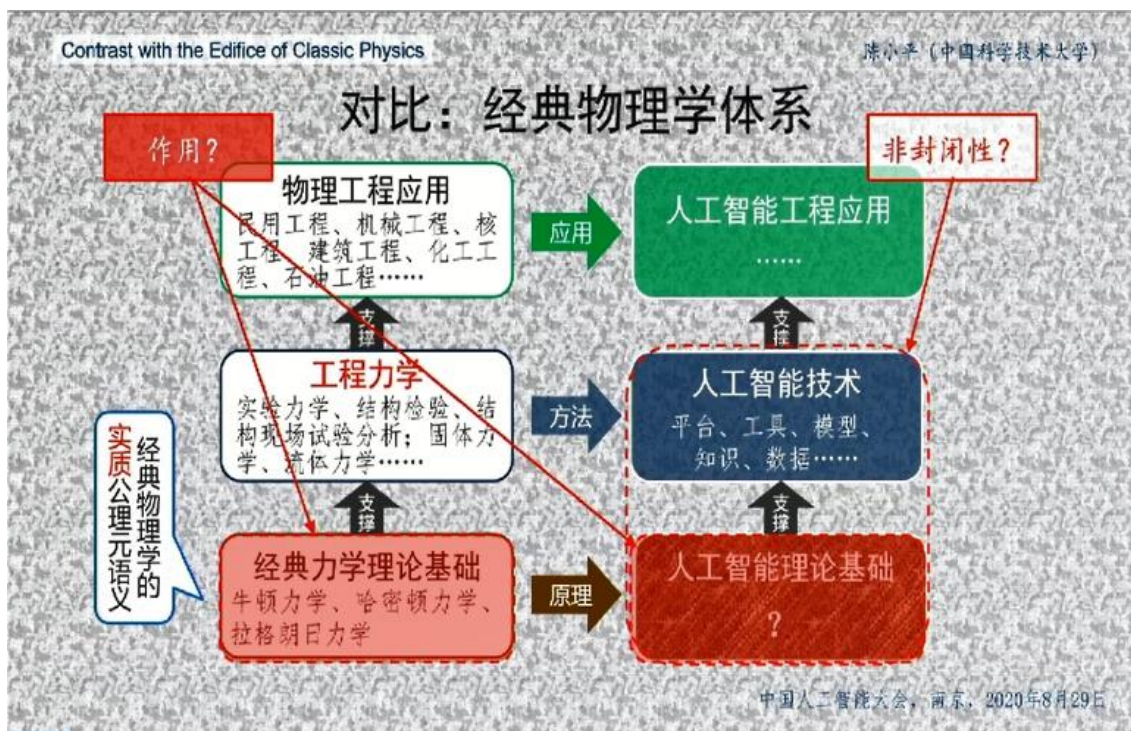
陈小平 (中国科学技术大学)

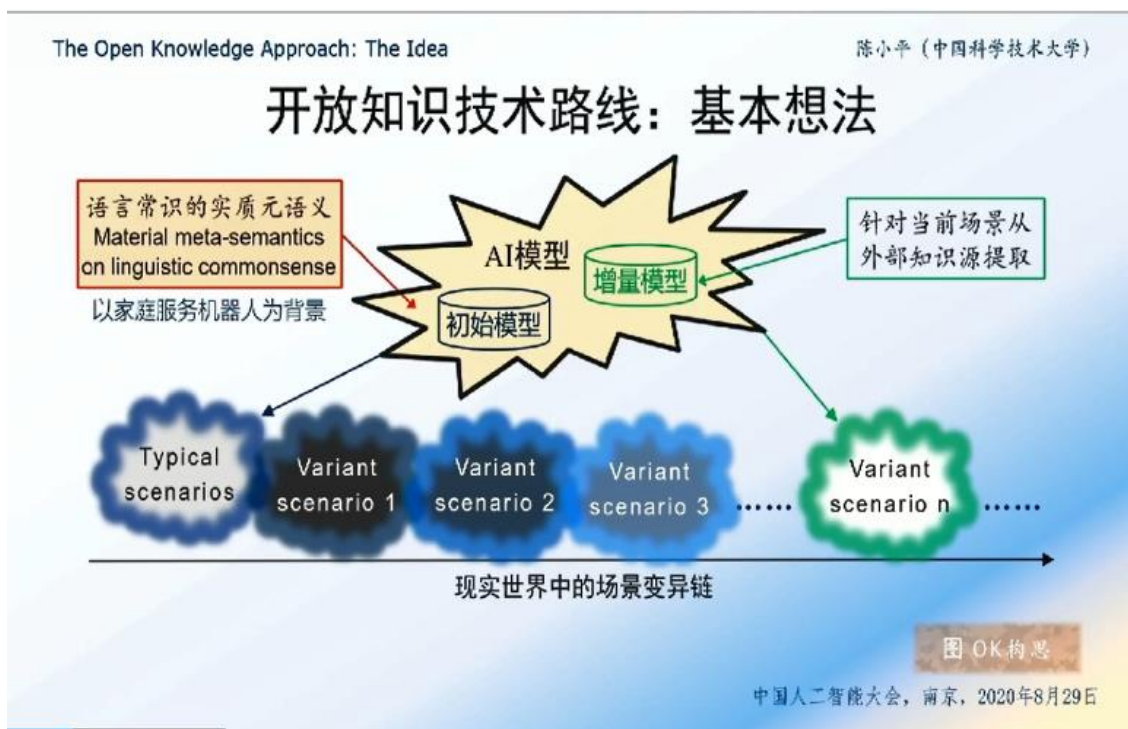
封闭性的定义与意义

- ❖ **定义** 如果场景S的降射 g 不是无尽的，S不含难解变元，并且S上的AI应用不存在致命性失误，则称S具有封闭性。
- ❖ **意义** 封闭性场景理论上可解；非封闭性场景面临三重挑战：
 1. **科学挑战**：AI的现有基础性数学工具不支持无尽性研究；
 2. **技术挑战**：难解变元的可解性没有理论保证，在可解情况下通常只有一事一议式的求解方法；
 3. **工程挑战**：致命性失误须通过场景裁剪/改造加以消除。

陈小平，人工智能中的封闭性和强封闭性——现有成果的能力边界、应用条件和伦理风险，《智能系统学报》2020年第1期/CAAI Trans. Intelligent Systems, 15(1): 114-120.

中国人工智能大会，南京，2020年8月29日



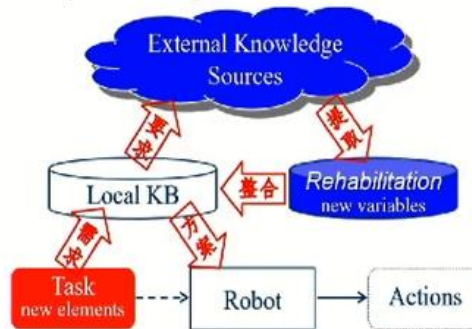


Formalization of the OK Framework in Causal Logic

陈小平 (中国科学技术大学)

形式化：知识修复问题

- ❖ A Knowledge Rehabilitation Problem (KRP) is a triple $\langle A, T, O \rangle$, where A and T are causal theories representing a local KB and external knowledge sources, and O a propositional formula representing a set of tasks for the robot.
- ❖ A causal theory E is a credulous rehabilitation for a KRP $P = \langle A, T, O \rangle$, if $E \subseteq T$ and there exists a model I of $LKB \cup E$ such that $I \models O$.
- ❖ The computational issue is how to find such an E effectively and efficiently.



Xiaoping Chen, Jianmin Ji, et al, Handling Open Knowledge for Service Robots, *Proceedings of IJCAI 13*, Beijing, China, Aug 3-9, 2013: 2459-2465.

Figure: The KRP

2019AI科学前沿大会，北京理工大学，2019年4月12日

Open Knowledge: Computational Complexities of KRPs

陈小平 (中国科学技术大学)

知识修复的计算复杂度

ECT $P = \langle A, T \rangle$	Credulous Rehabilitation		Knowledge Gap	
Problem:	Arbitrary	Definite	Arbitrary	Definite
Recognition	Σ_2^P	NP	Π_2^P	coNP
Consistency	Σ_2^P	NP	Σ_2^P	NP
Relevance	Σ_2^P	NP	Σ_3^P	Σ_2^P
Necessity	Π_2^P	coNP	Π_2^P	coNP

Each entry C represents completeness for the class C. The entries in the column under "Arbitrary" are complexity results of arbitrary extended causal theories and the entries under "Definite" are complexity results of definite causal theories. The entries in the row of "Recognition" are complexity results for the problem of determining whether a set $E \subseteq T$ is in the corresponding class.

Jianmin Ji and Xiaoping Chen, A Weighted Causal Theory for Acquiring and Utilizing Open Knowledge, *International Journal of Approximate Reasoning (IJAR)* 55(9): 2071-2082, 2014.

2019AI科学前沿大会，北京理工大学，2019年4月12日

Open Knowledge: A Tractable Case 陈小平 (中国科学技术大学)

A Tractable Case

- ❖ A KRP $P = \langle A, T, O \rangle$ is called *regular* if there exist sets S and J of atoms such that:
 - i) for each atom $a \in S$, a does not occur in A ;
 - ii) each causal law in T is of the form

$$a_1 \wedge \dots \wedge a_n \Rightarrow a \quad (2)$$
 where $a \in S$ and $a_1, \dots, a_n$ belong to $S \cup J$;
 - iii) O is a conjunction of atoms in S ;
 - iv) for any subset $J' \subseteq J$, there exists a model for the causal theory $A \cup \{ \neg a \Rightarrow \perp \mid a \in J' \}$.
- ❖ A credulous rehabilitation for a regular P can be computed in $O(n^2)$ time, where n is the number of causal laws in T .
- ❖ The rules in *Tasks/Steps* and *Help* of OMICS are of form (2).

Xiaoping Chen, Jianmin Ji, et al, Handling Open Knowledge for Service Robots, *Proceedings of IJCAI 13*, Beijing, China, Aug 3-9, 2013, pp2459-2465.

2019AI科学前沿大会, 北京理工大学, 2019年4月12日

OK: Experiments using qualified on-line common-sense knowledge 陈小平 (中国科学技术大学)

实验：机器人基本能力+预选的线上知识体

- ❖ **Open knowledge** from OMICS and a semantic dictionary:
 1. Procedural knowledge from the OMICS table *Tasks/Steps*
 2. Functional knowledge from the OMICS table *Help*, where a desire in *Desires* is mapped into some tasks (only in test 2)
 3. Conceptual knowledge (of synonym) from *WordNet*
- ❖ **Action Models** (Local knowledge, a robot's physical ability):

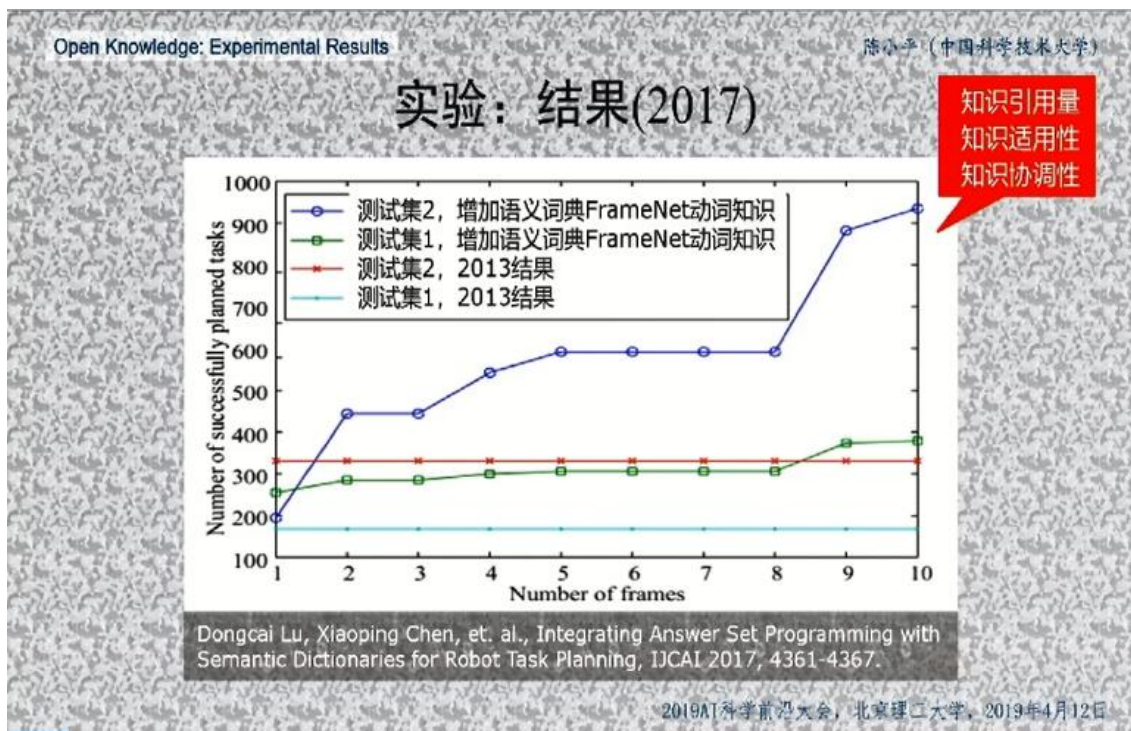
AM1 (*move*),

....

AM5 (*move, find, pick_up, put_down, open, close*)
- ❖ **Test set 1:** 11,615 user tasks from the OMICS table *Tasks/Steps* (e.g., "get food from refrigerator.")
- ❖ **Test set 2:** 467 user desires from the OMICS table *Desires* (e.g., "I am thirsty.")

Xiaoping Chen, Jiongkun Xie, Jianmin Ji, and Zhiqiang Sui, Toward Open Knowledge Enabling for Human-Robot Interaction, *Journal of Human-Robot Interaction*, 2012, 1(2): 100-117.

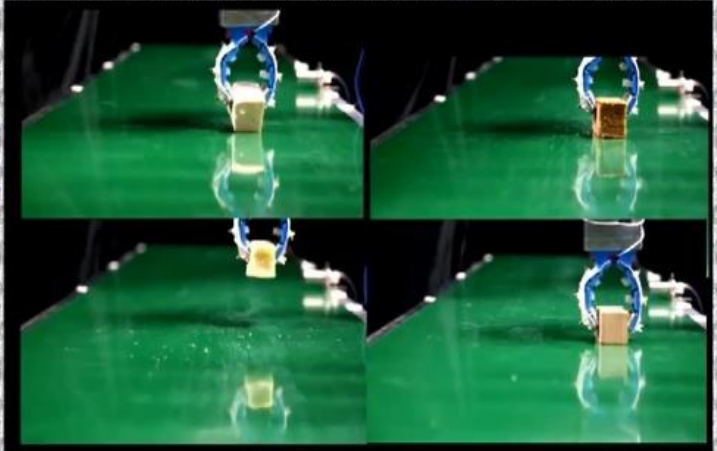
2019AI科学前沿大会, 北京理工大学, 2019年4月12日





OK Ability: Coping with Unexpected Scenario Elements 陈小平 (中国科学技术大学)

OK体系的融差能力：应对对象变异



不改程序和硬件参数，不用力传感器，操纵不同刚度物体。

清华大学，2019年9月27日

Future Work in the OK approach 陈小平 (中国科学技术大学)

OK的未来工作

- ❖ 涉及降射的因果推理 Causal reasoning involving grounding to various real-world scenarios
- ❖ 异质知识系统性整合的原理和方法 Principles and techniques of combining heterogeneous knowledge systematically on the OK framework.
- ❖ 融差性的理论基础与系统性方法 Theoretical foundation and systematical techniques of the principles of Rong Cha.

中国人工智能大会，南京，2020年8月29日

Prospect on 2020-2035: Extensive Applications, More Basic Research and AI Ethics Challenges Are Coming

陈小平 (中国科学技术大学)

展望：2020-2035

- ❖ 大规模产业应用 现有AI技术将在封闭性场景中得到大规模产业应用，优势领域包括制造业、智慧农业、交通运输业、部分服务业和信息产业，一大批应用企业正在崛起。
- ❖ 基础研究 AI发展模式正在从聚焦技术、基础薄弱、应用迟缓的状况向应用驱动、强化基础、深化技术的方向快速转变。
- ❖ 人工智能伦理 AI伦理正在成为人工智能应用和研究新的核心竞争力，带来巨大挑战和全新机遇。

中国人工智能大会，南京，2020年8月29日

图灵假说70年：两类AI与封闭性挑战

谢谢！

图片来自网络

陈小平 (中国科学技术大学)：中国人工智能大会，南京，2020年8月29日

相关推荐：

[CCAI 2020 大会特邀报告](#)

[CCAI 2020 大会特邀报告\(II\): Thoughtful Artificial Intelligence](#)

[CCAI 2020 大会特邀报告\(II\) , 微软亚洲研究院：多语种和多模态任务的预训练模型](#)

[CCAI 2020 大会特邀报告\(II\) , 华为：通用视觉：探索 · 实践 · 沉淀](#)



临菲信息技术港



临菲信息技术港公众号



临菲学堂