深度学习讲义

以数字速度做科学：AI 时代的科学未来

Nobel Prize Dialogue London 2026 深度讲义

Demis Hassabis 与 Paul Nurse 对谈

cnfjlhj & Pi

2026 年 9 月 4 日

来源类型：诺贝尔奖对话（公众科学对谈，中文配音版）
作者/频道/节目：Nobel Prize 官方频道 · Bilibili 中配 UP「旁白 _B」
发布日期：原片 2026-06-05（YouTube）· 中配 2026-06-11（Bilibili）
时长：约 88 分钟（原声）/ 86 分钟（中配 P1）
来源链接：`https://www.bilibili.com/video/BV1HuE66PESw`
原片链接：`youtube.com/watch?v=zVrA-_WP9Mw`

目录

1 开篇：当 AI「不再是同一个游戏」

2016 年 3 月，首尔。世界围棋冠军李世石（Lee Sedol）在五番棋中以 1:4 败给 DeepMind 的 AlphaGo。赛后他说了一句被反复引用的话：「这不再是同一个游戏了（It's not the same game anymore）」，随后从职业围棋退役。

十年后，2026 年 6 月，伦敦皇家学会（Royal Society）。Demis Hassabis——AlphaGo 与 AlphaFold 缔造者、2024 年诺贝尔化学奖得主——刚刚从韩国参加完这场比赛的十周年纪念回来，在那场纪念活动上重新见到了李世石。在皇家学会的舞台上，主持人 Adam Smith 把李世石那句话抛了出来，作为整场对话的收尾之问：**当 AI 被接入科学，科学还是同一个游戏吗？**

这场由皇家学会与诺贝尔基金会合办的「诺贝尔奖对话伦敦站」（Nobel Prize Dialogue London 2026），主题正是 *The Future of Science With AI*。台上坐着四位科学家，背景是墙上前皇家学会会长、诺贝尔奖得主 Rutherford 与 Florey 的画像。从李世石的「游戏之变」出发，对话在 88 分钟里穿越了：AI 如何改变科学发现的速度与方式、什么样的科学问题适合交给 AI、虚拟细胞与药物发现、创造力与「问对问题」、算力与全球不平等、AGI 风险与意识、科学哲学的新黄金时代，以及——科学这件事的根本是否被动摇。

本讲义按对话脉络逐段复原这场讨论，并在关键处补入机制性解释与教学图。读者应能在不观看原片的情况下，恢复这场对话几乎全部重要的论证转折、例证与保留态度。

出场人物

- **Julie Maxton**——皇家学会执行主任，开场致辞。
- **Hanna Stjärne**——诺贝尔基金会执行主任，欢迎辞（自动字幕将其名误识为「Hannah Graynee」，本讲义依官方信息更正）。
- **Adam Smith**——诺贝尔奖 Outreach 主持人，整场对话的主持与提问者。
- **Alison Noble**——牛津大学生物医学工程教授、皇家学会外事秘书，主持皇家学会《Science in the age of AI》报告工作组；研究领域为医学影像 AI（含超声与多模态数据）。
- **Demis Hassabis**——Google DeepMind 与 Isomorphic Labs 联合创始人兼 CEO，2024 年诺贝尔化学奖得主（AlphaFold 蛋白质结构预测）。
- **Paul Nurse**——弗朗西斯 · 克里克研究所（Francis Crick Institute）所长、皇家学会前会长，2001 年诺贝尔生理学或医学奖得主（细胞周期调控）。Hassabis 视其为自己在生物学上的「导师」。

2 两个开场：诺奖与 AI 的 2024 交汇

2.1 皇家学会的诺贝尔渊源（Julie Maxton，00:00–03:20）

Maxton 的开场是一次简短而精致的「机构史定位」。她提醒听众：从诺贝尔科学奖设立之初，皇家学会的院士就是最早的得主——荷兰化学家 van't Hoff 获 1901 年化学奖（渗透压工作），而他四年前刚被选为皇家学会院士；1902 年皇家学会院士更是在各科学领域「包揽」：生理学或医学奖的 Ronald Ross（疟疾研究）、化学奖的 Emil Fischer、物理学奖的 Lorentz 与 Zeeman（磁与辐射）。首位获诺奖的皇家学会会长是 1904 年物理学奖的 Lord Rayleigh；首位获诺奖的女性院士是 1964 年的 X 射线晶体学家 Dorothy Hodgkin。学会甚至「顺手」贡献了两位诺贝尔文学奖得主（丘吉尔、

罗素）与三位和平奖得主（Pauling、Borlaug、Rotblat）。

这段开场并非闲笔：它把当晚台上的三位——Alison Noble、Paul Nurse（现任会长）、Demis Hassabis——放进一条「皇家学会 = 诺贝尔科学传统」的连续谱里，为随后「AI 如何重写这条传统」定下张力。

2.2 2024：AI 一年拿了两个诺贝尔奖（Hanna Stjärne，03:20–07:02）

Stjärne 的欢迎辞点出了这场对话的现实坐标：**2024 年，两项诺贝尔奖都与 AI 相关**——物理学奖表彰「使机器学习以人工神经网络成为可能的发明」，化学奖表彰蛋白质研究突破（含用 AI 预测复杂蛋白质结构），而 Hassabis 正是当年的化学奖得主之一。她的潜台词是：AI 已经从「科学的外部工具」变成「被科学最高权威正式认可的科学内核」。这与 Alfred Nobel「以知识造福人类」的 125 年传统对接，也直接引出了当晚要追问的问题——AI 如何贡献于全球挑战、如何改造科学、而科学界又如何在善用 AI 的同时清醒面对它带来的挑战。

3 公众认知与皇家学会报告(Adam Smith × Alison Noble,07:02–19:45)

3.1 一份两年前的报告，今天读来如何

Adam Smith 把话筒交给 Alison Noble，请她回顾自己主持的皇家学会报告。Noble 透露了一个耐人寻味的细节：当初科学政策工作要选一个「颠覆性技术」领域来研究时，她曾半开玩笑地说「别回头又说是 AI」，但走出去和科学家们谈了一圈，结论还是 AI——因为它对「科学过程与方法论本身、以及科学家到底在做什么」的冲击，比任何可见的其它技术都更深。

报告成稿于约两年前（2024 年发布，调研覆盖此前 18 个月）。Noble 坦言，仅仅两年，公众对 AI 的理解已大幅跃迁：两年前人们还在问「怎么学 CNN」，如今因为用过 ChatGPT 之类的大语言模型，已经理解了一些大模型原理——但「个人对 AI 的看法」演变的速度，某种程度上比 *AI 本身的进步还快*。

3.2 报告浮出的关键议题

Noble 把报告里「至今仍高度相关」的几条列了出来：

- **算力可及性**（access to compute）——在英国内部已有推进，但仍是基础性门槛。
- **数据可及性**（access to data）——既是国内问题，也是国际问题。
- **可复现性**（reproducibility）——并非新概念，但在负责任地使用 AI 工具时面临新挑战。
- **技能与角色转变**——AI 往往需要多人承担不同角色，意味着从「单打独斗的计算学者」走向「团队科学」(team science)。团队科学对某些领域天然，对另一些则是全新事物，并连带出「该合作还是该竞争」的文化问题——而科学家天性爱竞争。

报告的著名设问也在这里出现：**AI 在科学中到底是「助手、同行还是导师」(assistant, peer, or tutor)**？这个三分法在当晚后半段被反复回扣。

3.3 推荐与国际差异

Noble 强调报告的建议高度因学科而异——不可能用一套伞形规则统管一切；新进入 AI 方法的领域应当向更成熟的领域学习（后者往往投资了高质量数据、设立了 data challenge，才准备好让机器学习成为发现新知的工具）。她自己的医学影像经历也勾勒出一条合作演进线：十年前是「计算方向临床要数据」；后来变成与终端用户（如超声技师）的「协同设计、协同创造」；如今部署系统后还要引入实验心理学，去理解人们「是否信任」这些工具、是否产生过度依赖。

国际差异一段尤为清醒：Noble 作为皇家学会外事秘书周游列国，发现态度并不一致——欧洲（含英国）强调「负责任的 AI」，将其作为科学的前提；而印度、中国对 AI 的潜能明显更兴奋，较少谈及顾虑，但也热切讨论伦理与责任问题。她特别提到在印度 AI 峰会上听到的一个关切：**大语言模型建立在「西方世界数据」之上，难以良好迁移到本土语言**——这是一个真实而具体的全球性议题。国家科学院因此成为跨越地理边界的、讨论这些问题的天然场所，而真正的壁垒往往是数据可及与信息共享。

3.4 报告之后

Noble 说，目前没有再出一部同规模报告的计划，但会持续追踪机器学习的演进——例如两年前报告里「提到但尚未泛滥」的生成式 AI，以及科学语境下机器学习模型「对不确定性的推理」（reasoning on uncertainty），都将是后续重点。「报告的目的本就是开启下一轮对话」，她以此作结。

本节要义

当晚的「前菜」其实已铺设了主菜的全部伏笔：算力/数据/可复现/团队科学、AI 的三重角色、国际不平等、以及「负责任」与「兴奋」之间的张力，都会在后半场被 Hassabis 与 Nurse 重新点燃并推向更深。

4 「以数字速度做科学」：Hassabis 的三层定义（19:45–27:50）

Hassabis 被 Adam Smith 请上台后，先回应自己那句被反复引用的口号——**science at digital speed**（以数字速度做科学）。他说这是 2020–2021 年看着 AlphaFold 折叠完全部约 2 亿个已知蛋白质时「冒出来的」。他把它拆成三层，第三层是「现在才真正显形」的新层。

4.1 第一层：解的速度本身

以 AlphaFold 为例：它不仅准确到「生物学家、实验科学家觉得有用」（虽不完美，但够用），而且几秒钟就能折叠一个蛋白质——极快。

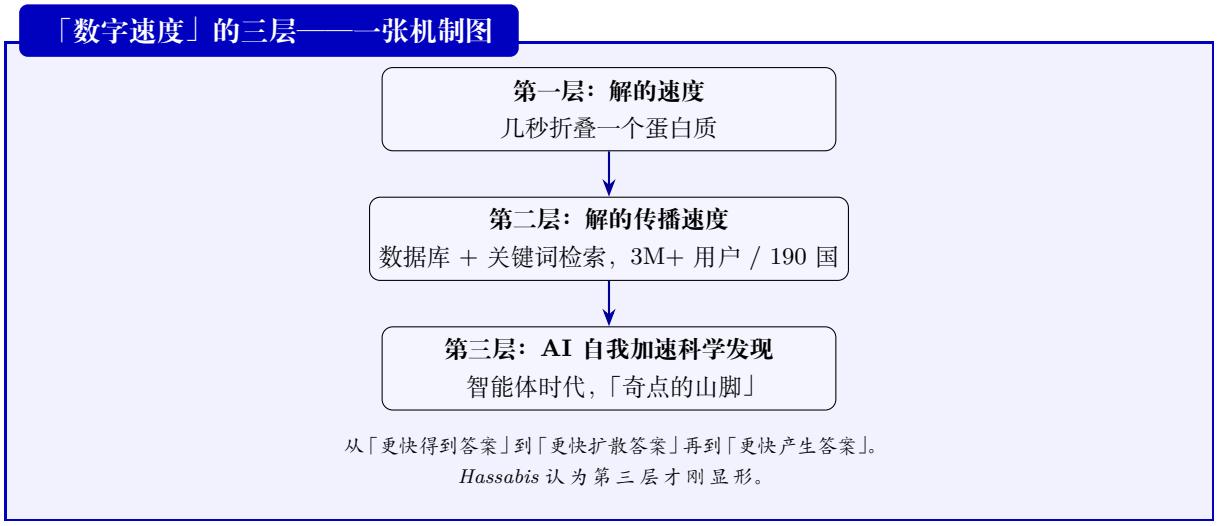
4.2 第二层：解的传播速度

Hassabis 认为这一层更接近「大科技的玩法」。一项新实验方法（哪怕是 CRISPR 这种量级的发现）从发表到渗透进湿实验室、成为博士生手里的标准工具，往往要「十年以上」。而 AlphaFold 一旦完成折叠，DeepMind 与 EMBL 下属的欧洲生物信息学研究所（EBI）合作，几个月内就把全部结构放进数据库，此后「简单到一次关键词搜索就能拿到」。如今全球已有超过 300 万研究者、来自 190 个国家使用过 AlphaFold 结构——「你甚至不需要湿实验室，无论身在何处都能访问并在其上继续建构」。他把这一层定性为「更像一件技术产品、一个应用」，是人们在科技业司空见惯、却在科学里头一次出现的东西。

4.3 第三层：AI 本身在加速科学发现

「而现在开始显形的第三层，是 AI 本身正在加速科学发现的节奏。」Hassabis 透露，他上周在 Google 年度活动上说了句引发骚动的话——「**我们正处在奇点的山脚（foothills of the singularity）**」，并强调「我是认真的」。他指的正是 Alison 提到的「智能体时代」（agentic era）：预测已久，但现在「第一次真正开始工作了」。回望十年，他认为这一刻（当下、明年）将被铭记为那场巨变的开端。

Adam 追问「奇点」这个词的分量。Hassabis 澄清：*AGI*（通用人工智能，由联合创始人 Shane Legg 命名）只是技术本身；「奇点」指的其实是围绕这项技术的更广泛的社会纪元——它将改写科学、经济乃至一切，而越来越多人在这一点上与他共识。



4.4 科学赶得上工程吗？

Adam 抛出一个克制的质疑：上周《Nature》刚发了两篇论文（其中一篇来自 Google DeepMind），AI 已经在「做实验、提假设」了。科学的强项之一本就是「需要时间来评估一件事是否真实、是否行得通」——现在还有这个时间吗？

Hassabis 的回答坦白而具张力：他「希望」有更多时间，能把科学方法更严格地用在 AI 上——理解我们正在构建的系统，它们不只是黑箱，要理解其深层运作细节；但「理解」落后于「性能」。而一旦 AI 变成有 chatbot、LLM 撑场的商业技术，「又给火添了柴」——**工程跑在科学前面**。作为「既是科学家又是工程师」的人，他希望科学能追上来，但承认这是当前市场力量的本性。话锋一转：即便如此，AGI「只剩几年」就要到，社会其实已经有过预警——AlphaFold 是 2020 年的事（如今看已是 AI 的「远古时代」），AlphaGo 距今整整十年（他刚从韩国庆祝这场比赛十周年回来），那场比赛「若回望，是现代 AI 纪元的起点」。「**这不是突然袭击，留意的人都有过 5–6 年的预警**」——问题只在于社会是否明智地用了这些年。

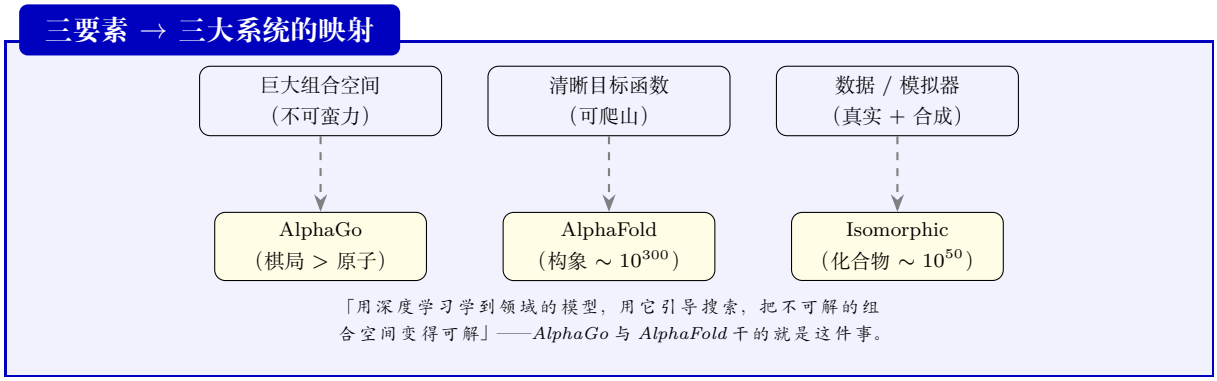
5 什么样的科学问题「可解」：三要素与 Isomorphic（27:50–33:10）

Adam 请 Hassabis 把「什么问题适合这套方法」讲清楚，并顺带问他兼任 CEO 的 Isomorphic Labs（雄心勃勃要「重造药物发现」，Nurse 是其顾问）。

Hassabis 说这套判断是 AlphaGo 之后渐渐成形的，可以用极一般的方式描述：

1. **巨大的组合空间**。围棋的合法棋局数比宇宙原子还多，无法暴力穷举；蛋白质构象数估计在 10^{300} 量级——他偏爱这类「不能用蛮力解」的问题空间。
2. **清晰的目标函数**。比如「赢下这盘棋」「找到最优一手」，或「最小化系统的自由能」——有了目标才能「爬山」逼近解。
3. **数据**。可以是真实数据，也可以是准确的模拟器，最好是两者皆有。这里有个有趣的交互：用现有数据造模拟器，再用模拟器生成额外合成数据。

他举了 AlphaFold 的合成数据策略做实例：PDB 里约 15 万个蛋白质结构「本身不够」，于是先用早期 AlphaFold 折叠上百万个蛋白，从中挑出 AlphaFold 自信「准确」的约 20–30 万个回填进去——同时强调**必须非常小心合成数据的分布是否匹配**。



药物发现正是这一框架的延续：Hassabis 估计「药物分子大小的化合物」约有 10^{50} 种，其中一种或数种可能契合某疾病的性质谱——问题在于能否找到、能否合成。Isomorphic 因此可以理解为「在开发另外半打 AlphaFold 量级的系统」：AlphaFold（蛋白结构）只是药物发现全流程里「一个小组件」，他们还要预测生物化学性质、毒性、ADMET（吸收/分布/代谢/排泄/毒性）等等。

6 湿实验室视角：从简单工具到闭环集成（Paul Nurse，33:10–38:45）

Nurse 自称「湿实验科学家」，坦率地说自己「绝非 AI 专家」。他给 AI 在自己实验室的贡献排了个由浅入深的序：

- **起步是「简单的事，但别看不起」**：机器学习让图像分析「过去要几天，现在几分钟」；几分钟做完一次媒体调研；大语言模型「给你的东西挺无聊，但能让你开个头」。
- **AlphaFold「确实惊人」**：做实验 → 设想可能的机制 → 过一遍 AlphaFold → 看什么可能触碰什么，「它或许对或许不对，但立刻就能生成想法与假设」。
- **下一步是「闭环」**：不是某个单一技术，而是把 AI 嵌进实验室工作的每一个阶段——把数据、假设、假设检验、生成新想法、再产出新数据这一循环连起来，让 AI 在每一环帮忙。

但 Nurse 也指出难点：AlphaFold 之所以成功，靠的是 PDB 那种「以完全相同方式收集、公开可用、巨量」的数据——「**要知道我们为什么做公共资助科学？没有它，AlphaFold 根本不可能**」。Hassabis 在一旁插话：「这也是我们之后把它开源的原因。」Nurse 紧接：「而且他们做得相当体面。」

问题在于，Nurse 认为未来最有用的生物学数据「大多不是那样收集的」：与其说是「大量同质数据」，不如说是「散落各处、深浅不一的井」，需要想办法把它们连起来。终极目标是「理解生命」，而起点是「理解细胞」。

Alison Noble 用超声的类比补了一刀：拿起超声探头扫描「机器人做不了，必须靠人类专长」，所以现在的研究方向是把一个个已解决的小任务「整合起来」——拿探头、移动、施加压力——其中部分能被建模，部分是「人类就是会做但难以言传」的细粒度信息，目前最难建模。「五年前这还是个梦，现在人们觉得可能了。」

本节小结

两位诺奖得主对「AI 已做到什么」的判断高度一致：从点状工具（图像分析、结构预测）走向全流程闭环（lab-in-the-loop），是当下正在发生的转变。但两人的清醒点也一致：AlphaFold 的成功建立在「公共、同质、巨量」的数据上，而生物学大部分数据并非如此——「散落各处的井」是下一道真实难关。

7 创造力、品味与「问对问题」(38:45–54:38)

Adam 把话题推向年轻科学家的焦虑：「当我建自己实验室时，世界会是什么样？创造性科学家的角色会是什么？」

7.1 把无聊的交给机器，把创造力留给人

Nurse 的回答带着一位老生物学家的真诚：「湿实验室研究极其无聊，真的无聊。我们大部分时间在把小量液体从一个管子移到另一个管子，没有任何办法能激发人的智力。」所以他主张：让机器人去做大量这类事，而**把焦点放到「如何让这些工具帮助人更有创造性」**。在他看来，大语言模型「在思想层面并不创造」，也许能把你推向某些方向；真正「点燃」科学家的是「创造性想法（大多当然是错的）」。他还怀旧地回忆 1970 年代「光是在显微镜下看东西」就有大量时间思考，并提醒一个反讽：用复杂技术时你会把全部时间耗在让该死的技术跑起来，反而没空思考你本想研究的生物过程——「但如果机器人替你做，也许我们又有时间想了。」

Hassabis 完全同意，并补上他当年「没进湿实验室」的原因正在于此。他的论点更系统：

- **问对问题比解决问题更难**。伟大科学家的「品味」体现在设定假设、挑选正确问题、把问题陈述得足够具体、设计好对照——「今天的 AI 系统完全做不到这些」。Nurse 在旁连声附和。Hassabis 补一句：并非永远不可能，但「未来 5–10 年大概做不到」。
- **更快地迭代「想法空间」**。像工程那样：快速搭原型、测试、推进。这会把工程式的迭代节奏带进更多科学学科。
- **一个博士生将能完成过去一整个实验室的量**。以他自己当年在 UCL 做 fMRI 神经成像为例：图像分析的苦活会被 AI 接走，省下的时间用于「想下一个好问题/假设」。他给年轻人的建议是「扎进去」——AI 不会回到盒子里，但理解它怎么工作仍重要（STEM、编程），因为「即使你以后不写那么多代码，理解了你能把编码工具用得好得多」。
- **工具会无处不在**。「像手机和 app 一样分布」——最有动力、最聪明的学生「无论身处哪个国家，都可能做最前沿的科学」，而不必再迁移到全球少数几个顶尖中心。

7.2 用技术「更少」，而非「更多」

Hassabis 抛出一个私人愿景：当 Gemini 这类聊天机器人/LLM「再过几年」变成真正能干的助手后，他梦想的是**让人「用技术更少，而不是更多」**。今天的困境是被「试图劫持我们注意力的不智

能系统」（社交媒体之流）轰炸，却又不得不钻进这条喧嚣之河去捞所需信息。若有一个可靠的、能代表你行事的 AI，「你可以说『我现在要去深度思考，10 点到下午 6 点别打扰我，到一天结束时把这个世界发生了什么、我关心的那些事汇总给我』」——以此保护我们的心智空间与注意空间。Nurse 打趣：「我都能想象我的研究生说要离开去『做深度思考』了。」

7.3 技术「被危险地诱惑」了（Nurse 的批评）

这是 Nurse 全场最犀利的一段。他警告：一些高影响因子期刊正在「被技术危险地诱惑」——「发明一种看问题的方式，堆一大堆数据，聪明、昂贵、只有少数人做得到，但**对任何有意义的事都下不了确定结论**，甚至看不出他们对下结论有兴趣——他们太沉浸于自己搞出了这个技术。」他点名《Cell》杂志「我读不下去了，全是这种东西，没有结论，因为他们根本不在想这意味着什么」。Nurse 的结论是：「**目标始终是理解，而不只是收集数据**」，而这是「人的问题，计算解决不了」。

Hassabis 同意，并举连接组学（connectomics）早期为反例——「尽可能多攒数据，却很少想它在功能上能告诉我们关于大脑的什么」。他把这放进一条更大的脉络：「大数据」是 2000 年代初的 buzzword，如今换成「AI」；他当年对早期投资人推销 DeepMind 的口径正是「**大数据是问题，AI 是解**」。

随后他给出一组重要判断：

- **AI 是生物学的「描述语言」，正如数学之于物理**。他认为细胞不会有「牛顿三定律」——「太涌现、太动态」，不代表没有规律可发现，但更可能是「一个模拟」的形态——这正是他们想做虚拟细胞的原因。
- 这套思路可推广到经济学等「以人为原子单元」的复杂系统，且这些领域无法写成优雅的数学方程。
- **已有成功先例**：AlphaFold 之外，DeepMind 在飓风路径预测上——英国气象局（Met Office）在用——过去要超算跑两周的流体动力学计算，「现在一个模型一样准甚至更准，几小时出结果」，对「向一个国家发出飓风预警」是天壤之别（去年飓风 Melissa 他们就做过）。
- 但他也承认「这些工具大多还用得很平常」（填表、看图），「那是第一步，该做」，而「**如何把工具用在你的问题上」本身就是一个创造过程**——连 CEO 们都常犯「为了用 AI 而用 AI，其实用统计就够了」的错。这又回到「品味」：既懂领域又懂机器学习，或找懂的学生，把工具正确地塑形到问题上。

Nurse 的警告（来源事实）

「技术的诱惑」这一批评出自 Nurse 在台上的原话，针对的是他观察到的「堆数据、无结论」的发表风气（他点名《Cell》）。这是他对当下科研文化的个人判断，本讲义如实复原，不将其上升为对所有数据驱动研究的否定——Hassabis 随后也区分了「有目标的大数据」与「无目标的大数据」。

8 虚拟细胞：边界、粒度与「适度的疯狂」（54:38–64:15）

Adam 把话题拉到两人「聊了 30 年」的共同梦想——**虚拟细胞**。

8.1 Nurse：微分方程路线「近乎无稽」

Nurse 没有客套。他认为目前常见的「简化路线」——取一个细菌细胞、减到 400–500 个基因、识别所有反应、建 400 个微分方程、再预测——「如果你有 400 个微分方程还什么都预测不了，那你不擅长微分方程」，并直言「开头这么搞近乎无稽」（但「该让人试，我可能错」）。关键障碍：这些动力学「几乎肯定是在体外测的，几乎肯定不会在体内那样工作」。

他随即给出一个更深、更哲学的框架来替代：

- 我们周围是「无序」，但细胞内是有序的。物理学家会因热力学第二定律皱眉，但一旦你解释「这是个隔离系统、有能量输入」，他们便失去兴趣——因为并不违反第二定律。
- 但 Nurse 强调：细胞不只是「造出秩序」，而是「**造出有目的的秩序**」——目的是让这个实体（细胞）存活、并在其上让自然选择起作用。「这难到凶残」，Nurse 自承「还看不太清怎么做」。
- 他目前最接近的想法（「我也不觉得是好主意，但最不坏」）：把细胞切成若干「域」（domain），当作**黑箱**——「不在乎箱子里在干嘛，只关心关键的输入与输出」。

Nurse 还讲了一个意味深长的实验：测量细胞内蛋白质合成的速率。同一培养物重复测量，均值完全一致；但看向群体，变异度高得惊人——大量细胞的速率只有正常水平的 50% 甚至更低。而生物学家的直觉是「一切都被严密调控」，Nurse 却指出：「其实非常松散、非常 sloppy。」更反直觉的是：钟形曲线尾端的异常细胞，10 分钟后又回到均值。他由此提出一个猜想——「如果调控太严，一旦你卡进控制空间的某个古怪角落，可能再也逃不出来；就像我们的电脑抽风时我们怎么办？关机再开机。细胞没法这么做。但如果它够松散，也许就永远不会卡死在错的地方。」他自嘲「这或许全是胡扯」，但坚持「像这样去想，再借助 Demis 正在做的东西」，是他会采用的进路。

8.2 Hassabis：从启发式到自学习，以及「超越人类知识」

Adam 把这串「松散即韧性」的思路，比作 Nurse 当年拿诺奖的工作——「把人类基因撒进本不分裂的酵母细胞」，这「有点疯狂，任何认真想的人大概都不会这么做，但它奏效了」。Nurse 笑答：「你得有点疯狂，也许模型能帮你『把疯狂的旋钮拧大』——调到恰好的疯狂度。」

Hassabis 接过话头，把这条线连到 AI 演进史：30 年前他想做虚拟细胞，大概会用「一堆微分方程加规则」的启发式方式——「那是当年 AI 的造法，比如国际象棋程序 Deep Blue」。但他在本科时已确信，这条路「走不到通用智能，也理解不了语言」——剑桥教的「一阶逻辑就是语言」显然不对，因为「我们常常不合语法，却照样互相听懂」。早期 DeepMind 的学习系统、再到 AlphaGo 证实了这一点：**围棋没法靠手写启发式做到世界冠军级**——「它太玄、太靠直觉、太关乎模式」，围棋每枚棋子价值相同，「国际象棋里好用的那些在这全不管用」，于是只能从经验与数据里直接学，靠自我对弈摸索出自己的「直觉启发」与有用策略。

为什么自学习系统能「超越人类知识」

AlphaGo 著名地发现了人类两千多年围棋史上从未出现过的策略。Hassabis 给出关键论证：一个被直接写进方程的系统（棋程序、或科学上的方程模型）很难超越其创造者已有的知识——「它哪里来的能力去超越?」而学习型、通用型系统则「有潜力在人类专家协助下，超越我们当前的知识水平」——这是人类造过的任何工具都从未具备的性质（「你的车不会突然自己飞起来」）。这正是他觉得这类机器「独特、令人着迷」的根本原因。

8.3 虚拟细胞：两个设计问题

Hassabis 把虚拟细胞归结为两个必须先回答的设计问题：

1. **划一个怎样的「边界系统」？**除非有朝一日把整颗行星量子层级地模拟下来，否则必须切出一个「自洽的、可近似外部一切的封闭系统」。
2. **以什么粒度建模内部，才对你的预测目标有价值？**

他举化学为例赞叹「物理学与生物学之间居然还有化学的位置」——化学可以「把量子物理抽象掉、又不必像生命那样复杂」，能独立研究、上下近似，这本身就很神奇，「也许科学的其它部分也能找到别的这种方式」。他透露自己当前的设想：**也许「细胞核」是一个好的起始子单元**（「大概还是太复杂」），正与合作者推进。

9 全球参与与算力（64:15–68:47）

Adam 问 Alison：Hassabis 说技术将「分布到全世界」，可她自己提过算力门槛——怎么确保「世界其它地方是参与，而不只是被裹挟」?

Alison 给出几条现实路径：如果工作依赖大型高性能计算，就换一种「不依赖大算力」的计算方式；提供算力可及性；用**联邦分析**（federated analysis）——把计算分散、再把结果汇总——既共享算力又保护数据本地性。具体取舍取决于问题与所需保真度。

Hassabis 更直言：算力问题背后是**能源成本**问题（英国电价全球最贵之列），而「能源现在基本上等于智能」——通过芯片与数据中心，二者将近乎一一对应。但他的核心忠告是「**这是个创造力与想象力问题**」：

- 建前沿模型要数百亿美元算力——「你在学术界或研究所，想都别想去干这个」，也跟不上、也非该干之事（美国有约 5 家、中国约 3 家公司在做）。
- **学术界真正该做的**是：分析、理解这些黑箱，压力测试其能力与极限，部署监控工具、做基准测试（benchmarks）——这些不需要大量算力。开源模型（如 DeepMind 的 Gemma 4，可跑在单台笔记本上；以及很好的小型中国模型，离前沿只差 6 个月到一年）让这一切触手可及。
- 他劝学界别被 FOMO 驱动：「最好的科学家，如果你信自己在做的事，就该像我们 2010 年那样屏蔽噪音——那时没人做 AI，所有人都觉得我们疯了。」而当下是「**做真正的多学科工作从未更容易的时刻**」：可用这些工具快速进入非本行的若干领域。「缺的，是创造性的想象。」

10 AGI 风险、对齐与意识（68:48–80:51）

10.1 Hassabis 的四层担忧

Adam 转入观众最关心的 AGI 与「意识」议题，问 Hassabis「什么让你担心」。

Hassabis 先重申自己毕生投身于此是为了推进科学、尤其是医学（AlphaFold、Isomorphic 即是他「亲手用这些工具」的体现），随后把担忧排成四层：

1. **恶意行为者**（bad actors）把通用技术挪用于伤害——从单个生物恐怖分子到流氓国家。
2. **AGI 风险（技术层）**：随着系统更自主、更强（智能体时代是第一步），我们能否保证护栏足以把行为约束在最初意图之内？「这是个超级难的问题。」

3. **经济问题**：即便解决了前两个，如何尽可能广泛地把利益分享给尽可能多的人与国家。

4. **哲学问题**：意义与目的。

他仍乐观（相信人类在「压力的关键时刻」会挺身而出），但惊讶于「我的经济学家朋友里认真对待、研究『后 AGI 经济体系』的人竟这么少」。

10.2 对齐：护栏能被「拔不掉」吗？

Adam 追问：能否造出「根本无法被移除」的护栏？Hassabis 称这是「大问题」——**对齐问题 (alignment problem)** 极其困难又极其有趣，且「最好在学术界或公民社会里做，而不是让大科技公司给自己的作业打分」。他「看着当下系统的可驾驭性」偏乐观，但「这是未解问题，谁觉得『找个开关就行了』，去读读相关文献」——Asimov 的全部机器人故事就是警告：三定律不工作，加第零定律也会被误读。

Adam 再问：发展这么快，有没有「带宽」做对齐？需要 slowdown 吗、长什么样？Hassabis 流露出一丝遗憾：他「20 年前想象的不是这个样子」——他原本希望基础研究能像 **CERN** 那样以国际协作、更科学的方式推进，那样应用仍可飞快（用 AlphaFold 之类在 AGI 到来前攻克癌症），而社会可以「准备好了再面对存在性问题」。但「语言机器人、聊天机器人变得能干又高度商业化，把这一切改变了，收不回盒子里了」。眼下只能另寻出路——围绕标准或认证做国际协作，而其难处在于：「正当我们最需要强有力的国际机构与协作时，我们却处在它的低谷」——地缘政治与机构碎片化是「糟糕的叠加」。他自承「有些想法」，但「这是根很难穿过的针」。

10.3 意识：可计算性，与两次「鲁比孔河」

Adam 直接问：人脑能做的一切，都可计算吗？

Hassabis 的偶像是图灵。他坦承与 Roger Penrose、Stuart Hameroff（主张脑中有量子过程）意见相左——「至今没看到证据」，并把那种论证戏称为「把两件我们都没搞懂的事凑在一起」。基于自己的神经科学工作，他倾向于「脑中大多数事、乃至一切，在极限意义上是可计算的」——而他造 AI 的原因之一，正是想把它当作神经科学的工具，也当作「对照/控制」来反观「心智到底有什么特别」。

Nurse 提醒一个前置困难：「我们其实不太知道意识是什么，所以连要解决的问题是什么都不太清楚。」Hassabis 同意，并给出关键区分：他们当年造「AGI」一词而非沿用「强 AI（strong AI）」，正因为「强 AI」把智能与意识搅在了一起，而他相信**二者是可分离的属性**——宠物猫狗「很多方面

挺有意识，但远没有人聪明」，提示意识可能是「渐变的、可分离的」。他断言「我们正在造的工具目前无论如何定义都谈不上有意识」，但「将来也许可能」。

他的建议是「**先只过一条鲁比孔河**」：先把 AGI（真正强大、精准的工具）做出来，给社会留出时间、并用这些工具更好地提出「意识是什么」、做神经科学（「给我十年，应该能做到」）；至于是否要过第二条河——*真的造出有意识的实体*——那是下一个该由社会决定的问题，而他倾向于「不该把两条河同时过」。

11 科学哲学与研究文化（76:44–83:03）

11.1 科学方法没有「金标准」（Nurse）

Adam 引观众提问：科学哲学在未来科学中扮演什么角色？Nurse 认为哲学「相当重要」——它关乎「什么是知识、什么可检验、什么不可检验」。他援引 Popper：你并不真的「证明」一件事对，只能证明它错，久而久之才「接受它也许对」，这里面缠绕着演绎与归纳之辨。他批评「我们没好好教『思考科学、如何做科学』这件有哲学基础的事」，但又反对「存在某种金标准科学方法」——「**在物理里怎么做科学，和在生物、气候学里怎么做，是不一样的**」。他点名批评「一些最伟大的物理学家在气候问题上毫无办法」，因为他们要求「从 A 到 Z 完全理解」才肯接受，而这套标准硬套到别的领域便会失灵。

11.2 「哲学的黄金时代」（Hassabis）

Hassabis 接得更远：「如果我现在是哲学家，这是有史以来最激动人心的时刻。」他认为需要一部*更新的科学哲学*：如何使用模拟、如何对待黑箱、如何把黑箱「逆向工程」回数学方程（「也许是两步走」）、「理解」现在到底意味着什么。而更广泛地，「我们需要新的康德、维特根斯坦、斯宾诺萨」——新的伦理哲学、德性、目的、人类境况的哲学，「如果其它技术问题都解决了，这些将是最重要的问题」。Nurse 打趣：「要是维特根斯坦，那我们就只能一直沉默了。」

11.3 研究文化：质量而非数量

Adam 念了一位年轻科学家的问题：AI 能否改善研究文化、尤其是年轻科学家承受的「骚扰」？Nurse 先给科研文化「降温」：「科学家和所有人一样会做坏事，但媒体爱把『都在造假』说成常态——其实很罕见，文化没那么糟，当然总有改进空间。」Alison Noble 则点出更普遍的焦虑：当下年轻科学家承受着「现在就要出结果」的压力，还担心自己的工作「上了 arXiv 可能已经被别人做了」。她的处方落在资深者身上：「**如果出版物和 CV 一直被当成头等大事，那就要由更资深的人站出来说：不，我要你做最好的科学家，不计发表数量**」——「几篇出色的论文，胜过堆量」。Nurse 一锤定音：「数量根本不重要，重要的是你产出的质量——这太显然了，竟有人不明白，真令人吃惊。」

12 收尾：还是同一个游戏吗？（83:04–87:55）

Adam 用李世石那句「不再是同一个游戏了」收尾，问四位：**接上 AI 之后，科学还是同一个游戏吗？**

12.1 Hassabis：苦甜与「游戏之谜」

Hassabis 讲了最近在韩国与李世石重逢的细节——「他状态很好」，正在以一种「切肤之痛」的方式体验 AI 侵入自己领域的感觉，「全世界现在大概都能从中学习」。他澄清一个被简化的叙事：李世石当时「已接近生涯末期」（Hassabis 形容他是「围棋界的费德勒」、18 次世界冠军、前十年最强棋手），「所以这些事被混在一起了」。

但这场重逢对他有「苦甜」之味——他自己年轻时曾是接近职业的棋手，「若非人生中某些不同际遇，我们也许会坐在棋盘对面」。他承认 AI 系统「确实改变了这种关系，正如国际象棋那样」；但随即话锋一转：「**国际象棋现在比以往任何时候都更受欢迎**」——「人们并不想看两台国际象棋电脑对弈，就像我们仍关心百米短跑与博尔特，尽管汽车比人快——我们想看同类人能做到什么。」围棋在亚洲「不止是游戏，是一种近乎神秘的艺术，被认为凝聚着宇宙的某些奥秘」；顶尖棋手曾告诉他「想知道棋的奥秘」——「可他们真的想知道吗?」Hassabis 把这对照回科学：「我有一种无法餍足的好奇心，也许病态地驱动了我一生——在场大多数科学家都有，否则不会做科学家。」我们在追寻自然与宇宙的规律、现实的本质，「AI 一定会帮上忙，但它会以某种代价到来。」

12.2 Nurse：工具不改变科学的根本

Nurse 的回答简短而坚定：AI「毫无疑问是个工具，能以不同方式帮你做科学——但那是正面的，不是负面的」。科学家的根本驱动是「理解我们此前不理解的东西」，动机是对未知的好奇。「坦率说，工具很重要，但我不认为它们改变了科学这一根本面向——理解世界。」

12.3 Hassabis：未来十年是「科学发现的黄金时代」

Hassabis 以一句乐观收尾：他很确信，「**未来 10 年我们将进入一个新的文艺复兴、一个科学发现的黄金时代**」；再往后或许还有别的纪元，但「至少未来 10 到 20 年我们会经历一段黄金时代」，而「回望时，AlphaFold 不过是我们借 AI 之力攻克的诸多例子之一」。

Adam 致谢，对话在皇家学会的画像与掌声中结束。

13 总结与延伸

13.1 贯穿全场的三条主线

回看 88 分钟，有四条彼此咬合的主线值得提炼：

1. 「**速度**」**的三重性，与**「**理解**」**的滞后**。Hassabis 的「数字速度」三层（解更快、传更快、产更快）是当晚的骨架；但他与 Nurse 都清醒于「工程跑在科学前面」「黑箱理解落后于性能」。这正是科学共同体的核心张力：AlphaFold 式的工程胜利已在改变生物学的日常，而对其为何奏效的科学理解仍在追赶。

2. 「**问对问题**」**是不可委派的核心**。两人从不同角度收敛到同一判断：AI 能折叠蛋白、能扫描文献、能跑闭环，但「设定假设、挑选问题、设计对照、判断什么值得问」这件「品味」之事，短期内仍只能由人承担。Nurse 对《Cell》式「堆数据无结论」的批评，正是这条主线的反面注脚。

3. 「**可解问题**」**的三要素是判断 AI 适用性的实用标尺**。巨大组合空间 + 清晰目标函数 + 数据/模拟器——这套判据把 AlphaGo、AlphaFold、Isomorphic 串成同一族问题，也为「什么时候其实用统计就够了」划出边界（Hassabis 对 CEO 们的劝诫）。

4. **从「工具」到「纪元」的跃迁及其代价**。Hassabis 把「奇点的山脚」「AGI 只剩几年」「未来十年黄金时代」连成一条社会纪元判断，并坦承其代价是对齐、经济分配、意义三道未解之题；Nurse 则始终把锚定在「科学的根本是理解未知，工具不改变这一性质」。

13.2 两人在「意识」与「方法」上的分歧与互补

Hassabis 倾向「脑中一切可计算」，并主张智能与意识可分离、应「分两次过河」；Nurse 则提醒「我们连意识是什么都没定义清」。这种互补恰是当晚最美的结构：一方提供技术愿景与时间表，一方提供生物学家的怀疑与对「目的性秩序」的敬畏。Nurse 关于「松散即韧性」的蛋白质合成实验、关于「细胞是有目的的秩序」的刻画，本质上是在提醒：*生命系统的目标函数本身尚未被形式化*——而这正是 Hassabis「三要素」里最难补全的「清晰目标函数」一环。换言之，虚拟细胞要真正成立，可能不缺算力，而缺一个关于「细胞在优化什么」的可计算表述。

13.3 延伸：与开放性 / AI-GA 叙事的呼应

这场对话的「AI 作为发现引擎」叙事，与你正在系统精读的 Jeff Clune 工作存在明显共振：Hassabis 的「AI 自我加速科学发现（第三层）」、AlphaGo「自我对弈发现人类未有的策略」，在机制上正是 Clune 所主张的*开放性*（open-endedness）与「AI 生成的 AI-GA」在具体科学场景中的落地样本；而「工具超越创造者的知识」这一点，也与 Clune 关于「搜索算法本身成为搜索对象」的论点同向。不同之处同样值得记下：Hassabis 的路径高度*目标驱动*（清晰目标函数 + 任务化搜索），而 Clune 的 AI-GA 愿景更强调*目标本身的开端性*——前者是「在已知问题里超越人类」，后者是「让系统能持续生成新的有意义问题」。Nurse 的「问对问题」警告，恰好站在两者的交界处。

13.4 留待追问的开放问题

- 若「清晰目标函数」是 AI 落地科学的门槛，那么细胞、意识、经济这类*目标函数本身未定*的系统，AI 该如何介入？「黑箱域」是否是一种妥协式的目标函数替代？
- 「科学赶得上工程吗」若答案是否定的，科学共同体能否承受一个「先用起来、后理解」的时代？可复现性与同行评议将如何改造？
- Hassabis 设想的「CERN 式国际基础研究」在当前地缘格局下是否还有可能？还是 AI 的对齐与治理注定要在碎片化中摸索？
- 「分两次过河」是一个价值判断还是一个技术判断？如果智能与意识*不可分*，这条建议便需要重新审视。

证据边界与来源说明

- 本讲义的**内容复原**基于该讲座官方英文原片的自动生成英文字幕（YouTube，Nobel Prize 官方频道，视频 `zVrA-_WP9Mw`，2507 条字幕，时长 87:51），经人工通读、去 ASR 噪声后按主题章节还原。用户提供的 Bilibili 中配版（BV1HuE66PESw，P1 配音 / P2 原声）为同一讲座；字幕因平台限制不可用，故以官方英文字幕为转录来源。
- 人名、机构、年份、奖项等事实，尽量依官方信息与对话本身核对；自动字幕中的明显识别错误（如将「Hanna Stjärne」识作「Hannah Graynee」）已据实更正。个别口语化、重复或语气词已在不改原意的前提下清理。

- 讲义中的**教学图**（「数字速度」三层、三要素映射、AGI 四层担忧）均为本讲义重绘示意图，用于梳理机制，并非讲座现场画面。
- 「延伸」一节中与 Jeff Clune 工作的对照，属本讲义作者的解释性综合，非讲座原文；讲座本身未提及 Clune 或 AI-GA。
- 封面取自 Bilibili 源视频封面；本讲义**未下载完整音视频**、**未抽取视频帧**（用户未要求「带视频」），遵循轻量素材原则。