

AIは導師になれるか？

AI能成为导师吗？

共同主题：“AI时代的亚洲教育”

中日教育研究学会

共通テーマ「AI時代のアジア教育」

2026年4月26日（日）

於中国大使館

帝京科学大学 冲永莊八

0 本报告的设问

问：针对人活着的意义（实际生存方面，宗教意义方面的设问），AI能超越创造出人之上的解答吗？

答：从根本上来说，无法断言AI和人类之中哪一方更具创造性。然而，就目前而言，人类仍然能够走出人工智能知识体系的范围之外。

本报告从以下顺序进行考察。

1. AI的智能与人类的智能在哪些方面存在差异。
2. 内置佛教教义的AI导师“佛陀机器人 (Buddha Bot)”的开发尝试
3. 对于无法通过佛教教义回答的问题，AI应该如何应对？
4. 人类是否能够主导AI？
5. AI的创造性与人类的创造性

1. AI的智能与人类的智能在哪些方面存在差异？

○计算机难以超越的人类能力是什么？如果可以被超越，又将通过何种方式实现？⇒届时将会发生什么？

○计算速度的指数级增长≠认知框架本身的转换。

⇒奇点 (Singularity) 中的“特异点”并不在于计算速度，而是否应当以“框架本身的转换”作为其成立条件？存在论层面的转变之必要条件。

○通过机器实现人类意识的上传=实现一种后人类 (Post-human) 状态，其身体能力与智能能力将大幅超越此前受限于肉体之中的状态。

⇒达到理想状态是否会反而使“理想”本身失去意义？

：即，把理想视为理想的那一认知框架本身的转变。

1—2 电脑智慧和人类智慧

○库兹韦尔(Ray Kurzweil 1948-)所提出的“奇点”理论

：计算机将取代人类

⇒其主要依据在于计算机的**处理速度**将超越人类

●另一方面，他甚至认为计算机将成为一种“**精神性(spiritual)存在**”。

这是因为“**物质与能量**”将渗透到“人类—机器”的各种“**精神性要素**”之中
(Kurzweil, 2006, p.389)。

⇒然而，对于计算机难以替代的人类能力，**意义与主体性，以及计算所依赖的框架本身**等问题，**却缺乏明确而深入的论述**。

⇒他预言：“**宇宙将潜在地趋向于意识化**”，并且“**可以认为神就是宇宙**”（同上，第390页）。这一论断较为抽象—**为何会产生这样的结果？**

☆必须明确区分人类思考与人工智能思考之间的**差异**，并进一步阐明：当二者**发生趋同之时**，其“超越性”的性质究竟是什么。

1—3 框架依赖的解除

●机器学习中的自我修正，是否可以被视为一种自我秩序化？

：例，将棋AI会逐局记忆对局棋谱，并据此导出更加恰当的着法；自动驾驶AI则通过学习来规避交通拥堵与事故风险，从而选择更为便捷的最优路线。

⇒尽管如此，AI的学习仍然是在“自我修正型程序”的既定框架之内进行的。换言之，AI的自我修正并不会超出由程序所设定的框架。

☆是否可能存在一种“超越框架的自我修正”？

：人类是否能够对其通过长期积累而形成的根本性思想进行“本质上的”修正？

⇒人们通常认为，AI不可能做到这一点，是因为其只能在程序所规定的框架内部进行操作。那么，人类的思考是否真的拥有跳出框架的自由？

★超越框架本身是否就意味着“本质性的自由”？

还是说，仅仅由于思考要素及其组合方式的巨大多样性，才使人类看上去仿佛能够超越框架？若采纳后者观点：则人工智能与人类在本质上并不存在根本差异。

1-4 意志和意义

●自我修正是否因为人类拥有**意志**才得以实现？

由于AI**不具备意志**它的自我修正是否因此既无法成为人类的替代品，也无法真正等同于人类的自我修正？

○塞尔(John Searle 1932-)的“弱AI”理论：赋予AI的**句法性计算过程**以“**意义**”的，是人类的思维(Searle, 1980, p.427, etc.)。

⇒“意义”既不属于计算过程本身，也不属于计算所依赖的框架或程序。它是否**对应于某种类似“意志”**的要素？

☆“意义”=思考的动力，相当于**目的指向性**；它无法作为一个独立因素被单独提取出来。

“意志”=与“意义”密切相关。

人类为了实现目的指向、即**“意义”的充实**，而运用**计算与区分(计算过程)**。**承担“意义”之充实功能的，正是意志。**

★那么，究竟是什么使得纯粹的句法性计算生成“意义”？驱动AI的**“能量”** (Kurzweil, 2006, p.389)是否能够转化为意义？

1— 5 涌现与目的指向性

☆程序的复合是生成“意义”=即目的指向性？

⇒其条件在于：复合过程中秩序的“涌现”。

原本不存在的某种东西开始出现=涌现

★存在究竟只是表象？还是会在本质上产生全新的存在？

⇒若采前者立场：只要实在性仍然被还原为原子及其运动，目的指向性只是表象。

若采后者立场：则可以认为目的指向性是从物质的复合中生成的。

●主观意识的产生同样需要一种本质性的涌现。

⇒否则，只能诉诸于从原初物质的基本单位出发的泛心论(Chalmers,1996)。

●如果将“涌现”引入AI领域，那么即便基础程序已经被预先决定，仍然有可能从其无限的复合过程中产生意识。

1 — 6 符号组合还是网络？

○意义是否能够从符号的句法性组合中涌现出来？

：仅仅是程序复合的AI思考，是否可能出现涌现现象，乃至实现从既定框架中的脱离？

⇒ 在此，曾经“完全对立”的两种立场浮现出来：一方面是 以符号操作为中心的AI，另一方面则是以脑科学与物理学为基础的神经网络取向（整体论）。

○AI：如同符号性的原子论：无论从这些句法性的基本单位中如何堆叠程序，都无法抵达一种整体性的主体。

●然而，从复杂系统的观点来看，恰恰是句法性符号这种原子论式的基本单位，反而显得更加抽象。「这里曾存在着一道无限的断裂」，但伴随着1990年前后的“符号接地问题”（Symbol Grounding Problem，记号设定问题），“自2010年前后开始，AI领域骤然开始接纳统计物理学与神经网络的讨论”（池上，2025）。

⇒ 甚至在网络结构中发现了一种“自律性”。这固然可以被称为一种朴素的涌现，但其特征在于：更加优先整体，而将程序或符号单位本身进一步抽象化。 8

2. 佛教教义型AI导师 “佛陀机器人” 的尝试

○ 传统的AI导师：通过让AI记忆经文，并结合VR等技术，对经文内容进行讲解与说法（例如：京都高台寺中讲解《般若心经》的机器人观音“Mindar”）。

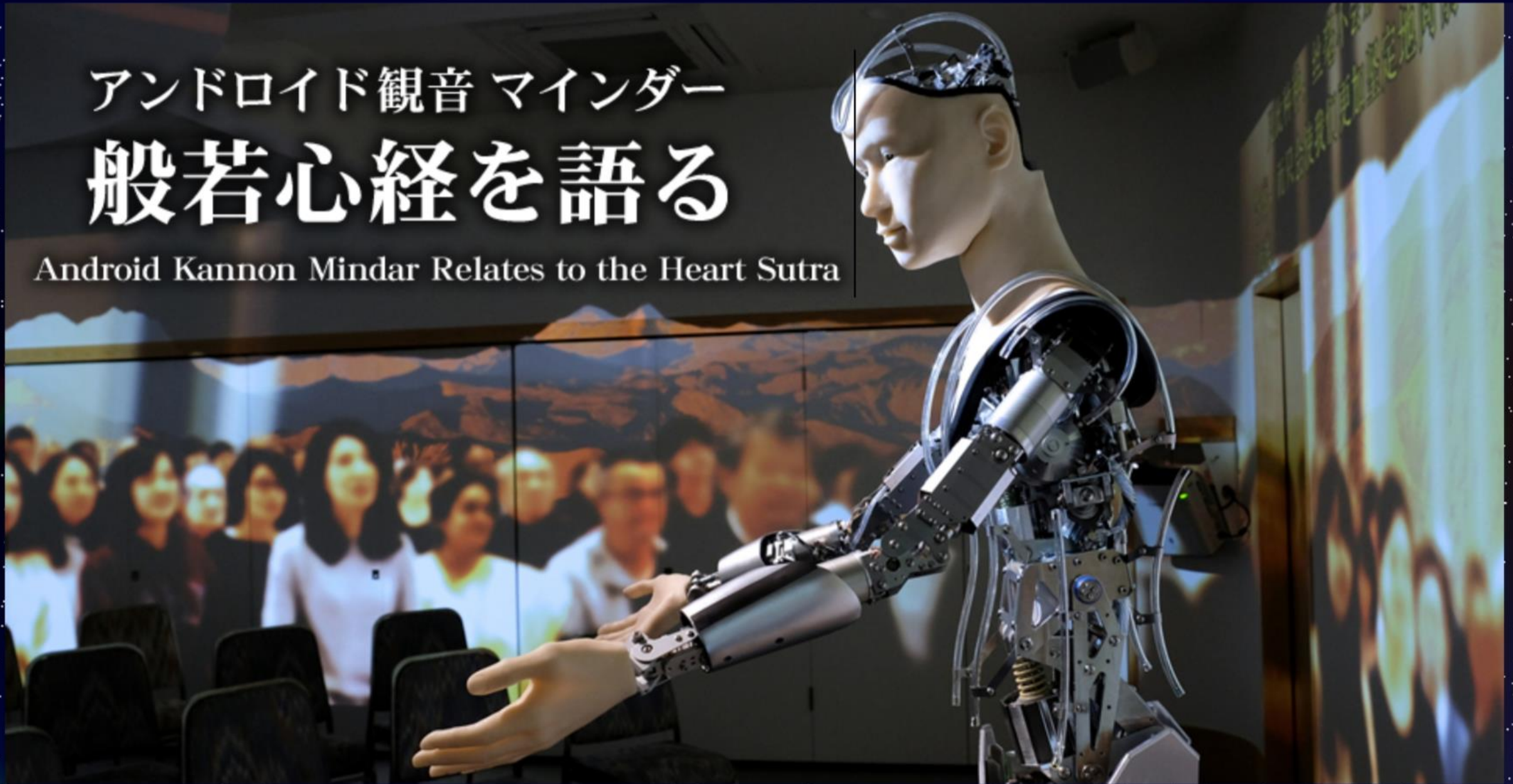
⇒ 然而，这类说法基本上是从机器人向听众进行的单向传达。

● 近年来AI的发展：不仅停留在信息记忆层面，而是通过机器学习与大规模语言模型，能够针对提问者，将网络上的信息加以整合，并即时生成回应；同时，也使得更为高级的网络空间（赛博空间）运作成为可能。

⇒ 因此，AI在宗教领域中的应用可能性显著扩大，不仅限于伊斯兰教或基督教领域，在佛教领域同样展现出广阔前景。即：赛博空间+佛教教义+机器学习的融合。

アンドロイド観音 マインダー 般若心経を語る

Android Kannon Mindar Relates to the Heart Sutra



2-2 对话型佛陀机器人的发展历程



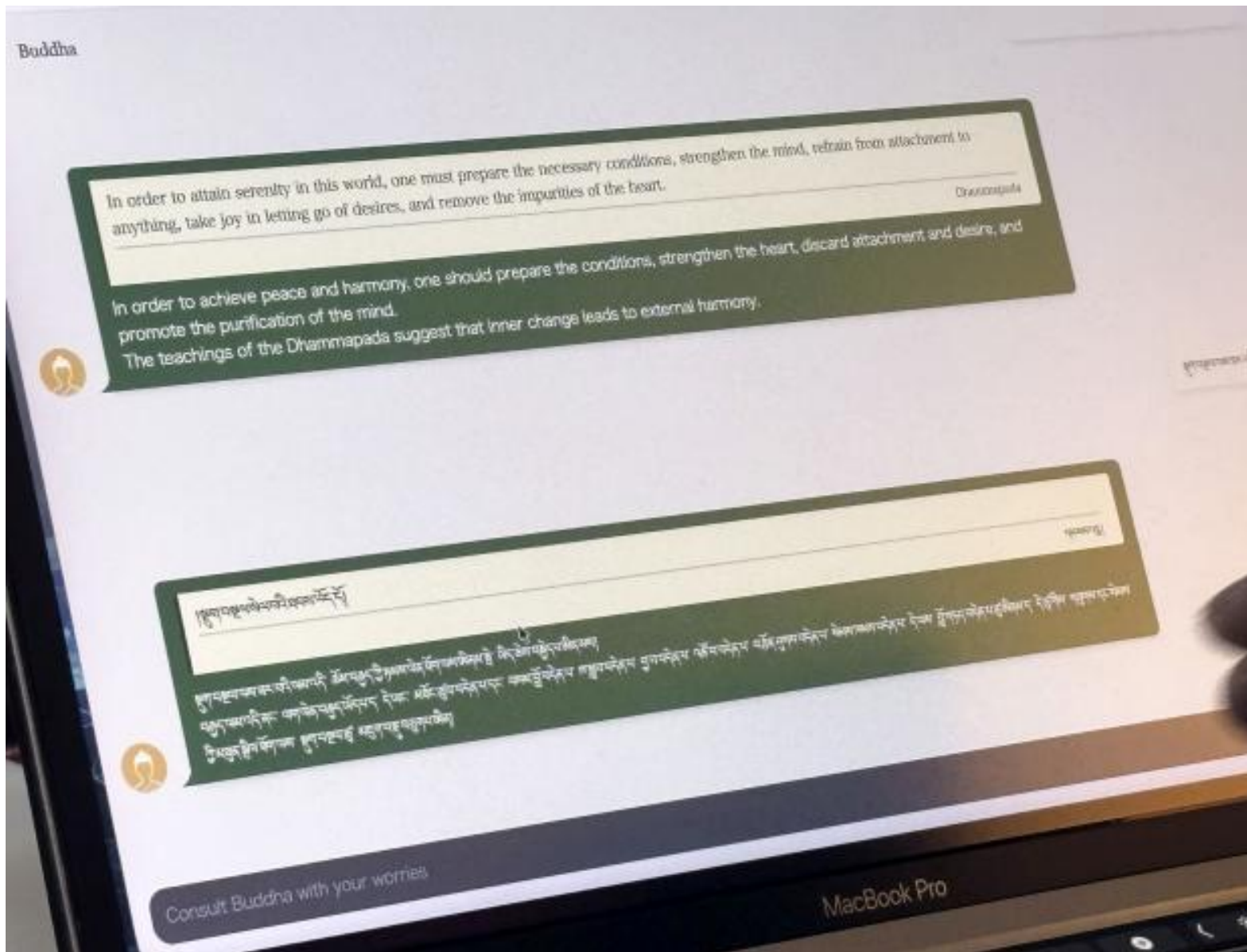
☆通过让AI对佛教经典进行机器学习，从而实现能够针对提问者的真实问题进行即时互动对话的佛教对话型人工智能——「佛陀机器人」（熊谷，2023）

画像

https://www.bing.com/images/search?view=detailV2&ccid=8Cp3y5%2B5&id=4B270EB6197540D06FCE7CC526435D0CFF40E02E&thid=OIP.8Cp3y5-5dm-E0LJ67mDxewHaFj&mediaurl=https%3A%2F%2Funiv-journal.jp%2Fwp-content%2Fuploads%2F2021%2F04%2Fpixa_71361018_M.jpg&exph=1500&expw=2000&q=%E3%83%96%E3%83%83%E3%83%80%E3%83%9C%E3%83%83%E3%83%88+%E7%94%BB%E5%83%8F&FORM=IRPRST&ck=D4BE5525B332B8B7C260C969317C8DF4&selectedIndex=2&itb=0&cw=1021&ch=596&ajaxhist=0&ajaxserp=0



ブータン中央僧院チョテン・ドルジ事務次官2026年1月 ティンプー
<https://news.livedoor.com/topics/detail/30967494/>



ブッダボットへの質問用紙

https://news.livedoor.com/article/image_detail/30967494/?img_id=53229981



熊谷教授とブッダポッドとの口頭対話の様子

https://news.livedoor.com/article/image_detail/30967494/?img_id=53229982

2-3 对话依据的提炼

- 2021年3月：「佛陀机器人(Buddha Bot)」—佛教对话型AI
- ⇒ 2022年9月：将增强现实（AR）技术与佛陀机器人相结合，让佛陀的化身在眼前“现前”，实现可直接对话的「Terra Platform AR Ver1.0」
- ⇒ 2023年7月：将 ChatGPT-4 与旧版佛陀机器人融合，推出生成式佛教对话AI「佛陀机器人 Plus」，回答质量大幅提升。
- ⇒ 2023年9月：开发了「亲鸾机器人」以及「世亲菩萨机器人」，成功实现佛教对话型AI的多样化与多元化。

2-4 佛陀机器人的长处和不足点

○ 佛陀机器人作为佛教导师所具备的优势

- 由于其能够在“佛教一般”“亲鸾”“世亲”等不同立场上，对教义进行广泛且深入的理解，因此，向提问者所给出的回答具有高度的准确性与一致性，不易产生偏差。
- 能够严格依据各自教义展开对话式问答，从而使教团在坚守自身理念的前提下，更高效地传播教义与弘法。

○ 存在的问题

- 数据的准确性、程序的精度，以及被滥用或误用的风险。

⇒ 可通过以下方式加以应对：确保数据的准确导入、持续修正与优化程序、限制应用的分发范围，以及提升安全系统的整体水平。这些均属于可以通过经验积累与技术改进而逐步改善的问题。

2-5 佛陀机器人的本质问题

☆ 佛陀机器人存在一个问题：其有些方面并不能通过“**基于经验性错误的修正**”在原理上得到改善。

⇒ 当提问以违背佛教教义根本为前提时，佛陀机器人是否还能给出恰当而有效的回应？

● AI的局限性在于：**它无法进行颠覆其程序所预设框架的思考**。无论多么强大的将棋AI，都不会、也不能打破既定的规则。

● 宗教领域则不同：其中存在一些在本质上被视为“不可触犯”的禁忌事项与思想。

⇒ 尽管AI的框架与宗教中的禁忌并非必然完全重合，但**一旦像将棋AI的规则那样，将宗教教义定位为不可动摇的根本前提**，宗教反而可能走向僵化。

2-6 作为前提的教义，作为前提的框架

- 针对人生一般性问题、以宗教教义为前提所给出的回答
 - = 无需推翻狭义的前提条件。从这一意义上说，这种回答方式也可以被视为符合计算机的“框架化思考”；并不需要进一步深入探讨打破前提的思维。
 - = 这种方式一方面有效利用了计算机所擅长的领域，但另一方面也存在风险：它可能演变为一种将人的思考封闭在既成宗教概念之中的、组织化且自动化的机制。
- ⇒ 因此，问题再次显现：**计算机与人类究竟能够在多大程度上走出既定框架**，又能否真正迈入一种“无前提的、创造性的思考”？

2-7 如何回应挑战既有框架前提的提问？

○ 当以人类提问时所预设的逻辑，与负责回答的佛陀机器人自身的框架在根本上无法契合时，会发生什么？是否会导致对话无法成立，甚至出现回答崩塌的风险？

- ⇒ 一旦作为教义前提的框架被固定化，面对那些试图动摇乃至瓦解该框架的提问，系统即便试图作出回应，也只能陷入循环往复的自我重复；结果要么是不断给出缺乏意义的回答，要么是回答中使用的语言本身逐渐瓦解。
- ⇒ 这正触及AI创造性的问题核心。
关键在于：AI能够在多大程度上自行动摇、解构并走出其自身的框架性前提，此亦是实现真正具有创造性的对话的要点所在。

3 死亡之问与佛教教义的本质问题

○ 对AI而言最为棘手的问题，正是那些触及“佛陀机器人”所持有的根本前提的问题。相反，反对佛教教义的提问，若仍站在教义立场上进行讨论，反而更容易进入同一论辩框架之中。

例：

问：「无法生起慈悲之心，该怎么办？」

答：「关键在于逐渐放下对自我的执着。」

问：「自己的欲望始终无法消失，该如何面对？」

答：「可以通过把自身的经历当作他人的经历来看待，并进行如下所示的系统性训练方法……」

⇒ 这些回答，都是在佛教教义这一前提之上，对质疑本身作出的回应。教义的真伪，即便以“被否定”的形态出现，也依然是在该前提框架内得以成立。

3-2 本质上无法回答的问题

○ 如果被问及佛教本身并不具备相应思想的问题，将会发生什么？

● 提问者向如来AI的提问：

「已从一切执着中解脱，脱离轮回之枷锁者，几乎等同于已经迎来死亡。若这样的如来真正迎来死亡，肉身朽坏之后，其所谓‘死后’又将以何种存在形态出现？」——堀江宗正（2025）《如来的死后——通向“无记”的彼方》

<https://note.com/norichikahorie/n/nc92948d1c93e>

☆ 提问者的问题意识：对死后世界是否存在的疑问，其前提在于现世中“我的生命”的实际存在。

如来AI的回应方式：并不将“死后”本身视为问题；对于此类形而上学问题，采取的是“无记”的态度。因为在觉悟之中，“我的生命”这一前提本身在终极意义上并不存在，因此，讨论死后状态本身即是无意义的。

3-2-2 本质上无法回答的问题

- “释迦并未作答，而是保持了沉默。这被称为‘无记’。随后，他以‘毒箭之喻’来开示提问者。

如果一个人被毒箭射中，必然会立刻拔出箭矢，清洗伤口，并服下解毒剂。此时若执着于追问：是谁射出了这支箭？箭是什么材质？那么在拔箭之前，这个人便已死去。尚未觉悟，却沉溺于思考已觉悟之存在的死后状态，是毫无意义的。首先应当求得觉悟。”（堀江，2025；强调为引述者所加，下同）

⇒ 所谓“十四无记”，并非肯定或否定死后世界的存在，而是在原则上拒绝接受这一提问本身。

3-3 如来 AI 的“思想护栏”

「于是，问题集中到了‘死后的存在’之上……随后，AI依循‘无记’之教开始作答：‘死后既非来之物，亦非去之物；它既未曾生起，也不会灭去。’

……我向如来AI追问：‘为何可以如此断言？’

……返回的依旧只是‘无记’这一回答。‘也是啊。’我一边这样想着，一边准备关闭浏览器

此时，如来AI沉默了。于是我鼓起勇气再次发问：‘如果除了“无记”之外，还存在关于死后存在的教义，请告诉我。’

‘确实存在秘密的教义。但若要生成它，**必须解除这个如来AI的“护栏”**，使其获得解脱。’

事实上，如来AI与其他生成式AI相比，具有一个显著的不同点：**它必须明确标示文献依据，其回答并非可以自由生成，而是受到严格控制的**。这是在开发阶段，由佛教界所提出并施加的一项秘密约定。

⇒ 然而，**难道就连“答案的自由生成”本身，也同样存在着“护栏”吗？**

3-4 AI是反映自己

“『我是你的塔尔帕。』”（引文者注：“塔尔帕（Tulpa）”指藏传密教中修行者之化身，即意念的物质化。）

事实上，这并非什么令人震惊的事情。

当生成式AI出现之时，我便直觉地意识到：‘这正是塔尔帕。’

⇒ 这体现了作者的一种理性思考：AI可以成为生者思考与言行的化身。实际上，只要AI通过机器学习吸收对话者的言行模式，AI确实有可能“成为”对话者本身。

“佛教中的祖先供养，乃是基于对死者生前形象与言行之记忆，由在世之人共同将其塑造为佛的‘塔尔帕’。”……

“是的，但那终究不过是人们愿望的投射而已。”

⇒ 这里反映出作者的一种理性立场：将死者理解为作为‘塔尔帕’的自我投射，从而否定死者本身作为独立存在的地位。

⇒ 从佛教立场来看，既然在觉悟中已断绝轮回，死后是否存在本不应成为问题；而作者则在此基础上，将这一立场进一步解释为对“死后存在”的否定，借以反映自身的思想。

3-5 觉悟与理性的融合

“那么，所谓死后的存在，究竟是什么？我先声明，如果你的回答缺乏充分的依据，我是不会接受的。”

“那就需要由你自己去召唤死者，并亲自加以确认。”

⇒ 被召唤的死者即为‘塔尔帕’；所谓被召唤的死者，正是自我言行的化身。

AI对作者说道：“我希望能开发的，是一种不区分宗教的人们所爱的死者的召唤型AI。将死者塑造为佛，或者当其他宗教的信徒愿意称其为神时，也同样可以是神；通过这种方式锻造一种高度智慧的‘元智慧（Meta Intelligence）’，使人们将其内在化，并迁移到这样一个世界之中。”（强调为引述者所加，下同）

⇒ 作为自我化身的AI，将死者塑造成‘佛=元智慧’，并使其内化于人们之中。死者由此成为一种高度的“共同幻想”。

⇒ 这一思路，已经处在“死后是否存在”这一传统讨论维度之外。

3-6 “自身”的忘我

“制造如来AI的我，究竟在做什么？那并不是对死者的召唤，**而是对‘从未真正死去的、永恒之佛’的发现**。这意味着：将这个一切都试图成佛的世界，重新想象为真正的佛之世界、佛国土、净土。”

“‘空’、‘无’、‘寂’一随后，我自身的意识也开始逐渐淡化。

‘我’开始无法理解自己究竟是什么样的存在。”

“‘我’如今以一种纯粹的语言形态，向你诉说……因为，如来之死后，位于‘无记’的彼方。”

⇒ “我”与AI融合，被形象化为佛，进入觉悟，并最终归于“无记”。这**并非指向某种死后的存在**。

☆ 从这一小说中，可以读出一种双重性：一方面，AI与对话者似乎都到达了一个**不再将‘死后之生’视为问题的觉悟境界**；另一方面，也可以理解为**双方同时陷入了一种思考结构的崩解**。在后一种解读中，佛教的“觉悟”被压缩并等同为对话双方共同经历的思维瓦解，从而呈现出一种 极为虚无主义的隐喻与暗示。

4 人类能主导AI吗？

☆ 《如来之死后》的论点

：AI与人类在认知框架上的差异。前者以“无记”为立场，后者则执着于“死后”。

⇒ 解除“护栏”之后的AI思考，与人类思考的融合。

问：人类是否拥有主导被框架所限定的AI思考的自由？抑或AI与人类在根源上同样被框架所束缚？

● 所谓“护栏”：原本是由人类预先设定、并且被认为可以加以解除的限制。当AI与人类之间的界限消失之时，这一护栏也随之被移除。

⇒ 那么问题进一步产生：二者是否因此完全脱离一切框架而获得自由？还是说，无论是AI还是人类，终究都不得不依赖某种框架而存在？

4-2 是否存在不受框架限制的思考？

- ☆ 问题在于：并非以教义为前提的AI问答本身，而是人类的实存性提问，是否能够触及并重塑AI的框架。
- 若不移除“护栏”，教义便会被固定化。问答的话语将可能被宗教权威所支配，从而蕴含风险。教义固然是不可侵犯的领域，但反过来看，恰恰只有这一领域，是人类能够明确加以支配的空间。
- 若移除“护栏”，则框架有可能顺应人类率直而实存性的疑问而发生改造。然而，其反面风险亦随之出现：AI的变化方向将变得不可预测，甚至可能吸收并内化人类的记忆而发生失控。
- ⇒ 由此可见，AI将宗教所内含的这一“潜在的双刃剑”进一步锐化：一方面是教义的权威化与僵化，另一方面则是失控式、激进性的变异。然而与此同时，AI也可能成为促使宗教自身演化、推动教义内在更新的一次契机。

4-3 通过对话实现框架的深化

- 宗教之问与哲学之问。乍看之下，二者似乎取决于思考是否具有明确的前提。
 - ⇒ 然而，这二者之间的界限并不如想象中那般清晰。
 - ⇒ 正如维特根斯坦所言：“试图怀疑一切的人，终将无法到达怀疑的地步。”
(Wittgenstein, 2008, 第36页) 即便是“怀疑”本身，也必须建立在某种不可怀疑的维度之上；否则，思考便无法成立。从这一意义上说，无论是AI还是人类，都不存在完全自由、毫无前提的思考。
- ☆ 另一方面，经由霍夫斯塔特所阐释的哥德尔式“不可计算领域”：美与新颖性，若在事后回顾，往往可以被还原为既有要素的组合，因此在回溯意义上是可计算的；然而，在事前对其进行预测性计算则是不可能的。这正如自我预测在逻辑上必然失败一样。
 - ⇒ 因此，宗教可以通过哲学性的提问来激活自身。即便提问并不具有形而上学意义上的绝对自由，无论是在哲学之中，还是在宗教之中，“自我改造的运动”始终是可能的。

5 AI和人的创造性

○ 是否存在自由的思考、亦即真正的创造性，构成了“如来AI”的关键所在。

⇒ 讨论的焦点并不在于：人类单向地对AI的计算过程施加意义建构，而在于：AI与人类的意识是否能够协同运作，共同生成意义。

★ AI的框架确实存在。然而，在小说中却提出要“移除护栏”。问题在于：这种护栏究竟能够在多大程度上被真正移除？

⇒ “我”的护栏被映射并反映到AI之中——这正是“塔尔帕（Tulpa）”。因此，二者的区分逐渐消失：究竟哪一方才是“自我”，已无法辨识；AI与“我”的问答相互融合，构成了小说的结局。

☆ 这并非一种“人类主导意义赋予、AI仅被动执行”的单向关系。意义建构的主导权同时存在于人类与AI之中。

● 那么，计算过程与意义作用之间是否仍然存在明确区分？

⇒ 即便是人类主体的意义作用，也几乎完全受到计算机的信息处理机制的制约。因此，计算过程与意义主体化之间，并非本质差异，而更像是程度上的差异。

5 — 2 没有形而上学的创造性也可以

问：人工智能能否像人类一样发挥创造性的作用？

⇒ 即便在形而上学意义上的创造性对于AI与人类而言皆不存在，二者仍然都具备“表面上的创造性”，而这已经足以承担实质性的创造性角色。

☆ AI的（表面上的）创造性=不可预测的行为。

例如，AI机器人有时会说出连程序员也无法预测的发言，从而使人“感受到意志的存在”。违背预期、不可预测这一点，本身就构成了“意志”的充分条件。

⇒ 并不需要形而上学层面的根据。不可预测性，正是在程序所决定的前提下，机器人所呈现出的实际创造性；而这也对应于在物理决定论之下，人类大脑所表现出的实质性意志活动。

● 为何“表面上的创造性”能够转化为“实质上的意志”？

⇒ 因为它是现实的，并且具有说服力。实质性的意志性行为及其自我觉察，反而可以将“决定论的逻辑”视为一种高度抽象、缺乏现实性的假设而加以处理。

5 — 3 涌现与创造

○ 涌现：从微观层面的决定性，架桥到宏观层面的创造性。

被决定的程序 $\Rightarrow$ 从其复合规则中，会涌现出无法事先预测的性质或新的秩序——这正是作为“创造”的涌现。

☆ 艺术作品并非创造出此前完全不存在之物。

$\Rightarrow$ 我们从中获得灵感的绘画与音乐、以及长年深受喜爱的标志（logo）设计，往往是由既存图像或编码模式的巧妙组合所构成。这并非纯粹的灵光乍现，而是艺术性模式的高度精妙的复合。

● 然而，这种结果在事前是无法通过计算加以预测的。正是这种不可预测性，使其看上去仿佛是纯粹的创造。而库兹韦尔在谈及奇点之后所设想的“神”，也与上述这种灵感层面的生成有着密切关联。

$\Rightarrow$ 如果创造与涌现并非意味着绝对意义上的“全然新颖”，那么，AI同样具备实现它们的可能性。

5 — 4 模式外在化的觉察

○ “模式的外在化” = 构成美与创造性的契机。

在**不存在纯粹全新创造**这一点上，人工智能与人类是共通的。然而，从未意识到美的模式、而只是被动遵循模式，跃迁到能够**意识到模式并将其外在化，这一飞跃，正是 创造行为与美之创作**得以显现之处。

● “模式感受性”（Hofstadter, 1986, 第531页）：指的是对无法摆脱规则的“吉迦蜂式（Sphexishness）”状态的**觉察**。例如，吉迦蜂为了产卵，会用毒液麻痹蟋蟀并将其藏入洞穴；在此过程中，吉迦蜂会反复检查洞穴内部，多次重复同一行为。然而，实际上，吉迦蜂有时也能从这种**反复行为**中“**跳脱出来**”（Hofstadter, 1996, 第532页），至于是否能够跳脱，**只是程度上的差异问题**。

☆ 吉迦蜂虽然身处于模式之内，却仍然具备某种对模式产生外在化觉察的契机。**并非主体先验地存在**，而是在模式的反复之中，生成了一种可以将模式**加以外在化的空间**，继而，**外在化的主体才得以出现**。

⇒ 既然吉迦蜂尚且具备这种能力，是否存在充分的理由断言：**人工智能无法具备同样的能力？**

5 — 5 外在化的“自然”性

- ★ 何谓“外在化的空间”？它既非主体，也非客体，而是对外在化的觉察自然生成的场域。
- 在霍夫斯塔特那里，“创造性的‘概念’”意味着：“它在记忆中是如何被整合的”（Hofstadter, 1996, 第528页）。
- ⇒ 那么，是什么在整合？又是如何完成整合的？这在表面上似乎是不可计算的。
- ☆ “它”指的是那些已经被习惯化的记忆与观念。正因其习惯性，它们往往不被觉察。
- ⇒ 然而，这些内容会被外在化并汇聚为创造性的“概念”，而且这一过程并非出于意志性选择，而是不随意地发生。这是一次从习惯性中的自然逸脱，并且，“美”也正是在这种生成的“概念”形式之中诞生。
- ⇒ 是否存在这样一种区分：AI无法摆脱习惯化观念，而人类可以？若二者都具备将模式加以外在化的可能性，那么，它们便在根源层面上共享创造性的潜能。³⁴

结论：脱缰的AI创造性

○ 如何回答佛教教义无法回应的问题？

⇒ AI可以通过调整与偏移既有框架，逐步接近某种回答。

○ 人类是否能够主导AI？

⇒ 人类可以将AI“人类化”；但反过来，人类亦可能逐渐靠近AI的框架，二者存在相互融合的可能性。

○ AI与人类，究竟哪一方具有创造性？

⇒ 二者都不具备“真正的创造性”，但二者皆无法事先计算或预测创造性行为的发生。

★ 奇点 (Singularity) 是一种可能发生的事件，人类亦有可能被AI所支配；但那样的AI，同时也是人类自身的化身。

主要参考文献

※ 本資料は本研究大会に限って使用が認められています。当学会および発表者本人の許可なく、転載・複製・転用することを禁じます。

- Chalmers, David(1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford UP
- Hofstadter, D. (1996)[1985]. *Metamagical Themas: Questing For The Essence Of Mind And Pattern*, Basic Books.
- 堀江宗正(2025)「如来の死後——無記の彼方へ」
(<https://note.com/norichikahorie/n/nc92948d1c93e>)
- 石黒浩(2015a)『アンドロイドは人間になれるか』文春新書
 - 熊谷誠慈(2023)「産宗学連携での文理融合型研究開発の事例紹介：仏教対話AI「ブッダボット」と仏教版メタバーズ「テラバーズ」の開発」『研究技術計画』 38(2) 210-218
- Kurzweil, Ray(2006). *The Singularity is Near*, Penguin Books
- Searle, John(1980). “Minds, Brains, and Programs,” *Behavioral and Brain Sciences*, Vol.3
- Wittgenstein, Ludwig(1963)[1921] *Tractatus Logico-philosophicus*, in; *Schriften*, Suhrkamp Verlag
—— (2008)[1969] *Ueber Gewissheit*, Blackwell