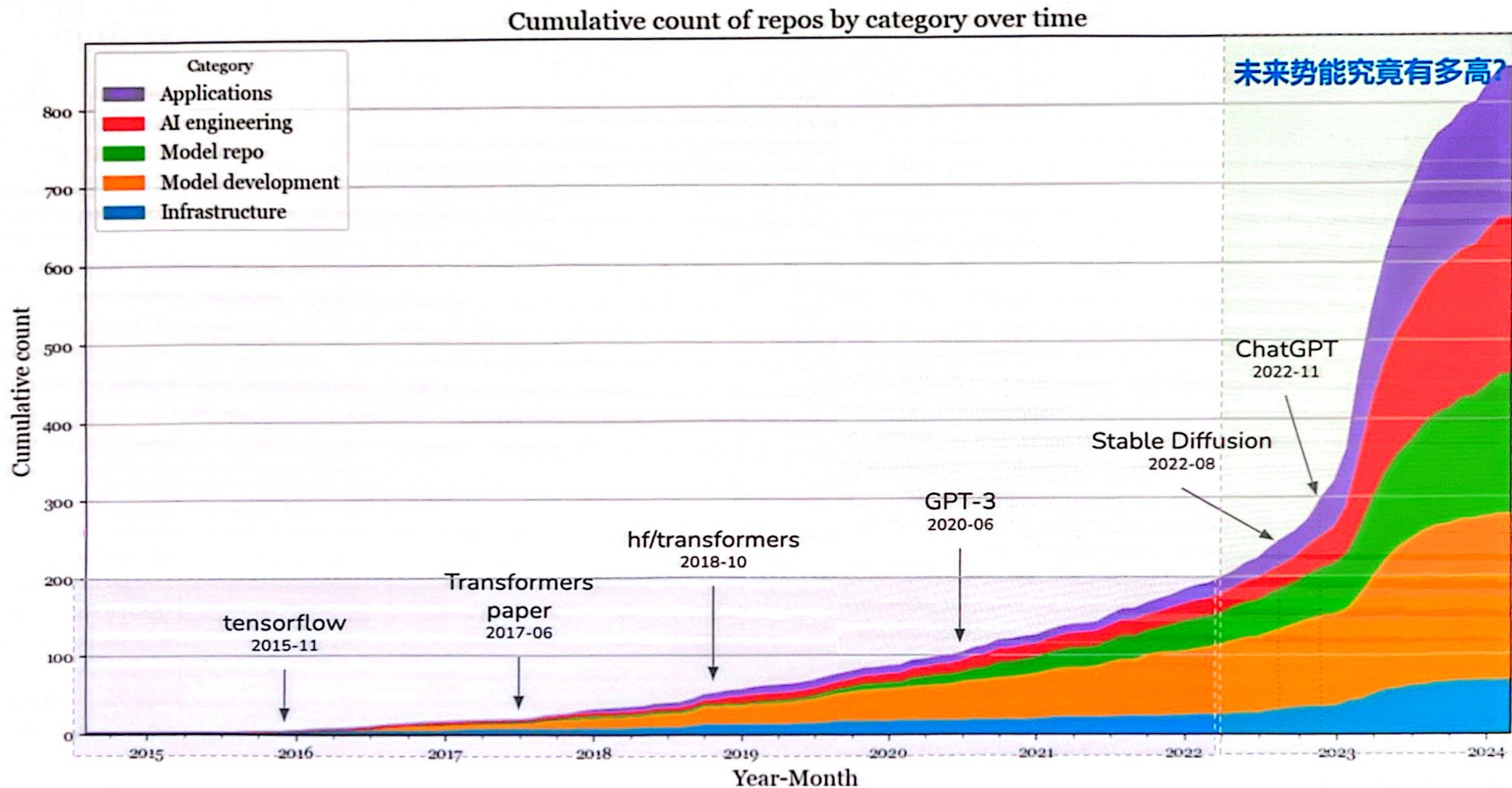


1、大势观澜：大语言模型的开源技术正在蓄力更大的技术突破



Linux操作系统、互联网、安卓系统、移动互联网、工业软件的发展与繁荣，无不是开源技术参与下从量变到质变的发展过程。

2、大势观澜：从10万、30万、70万到100万



Hugging Face
开源大语言模型托管平台

Model Growth Monthly



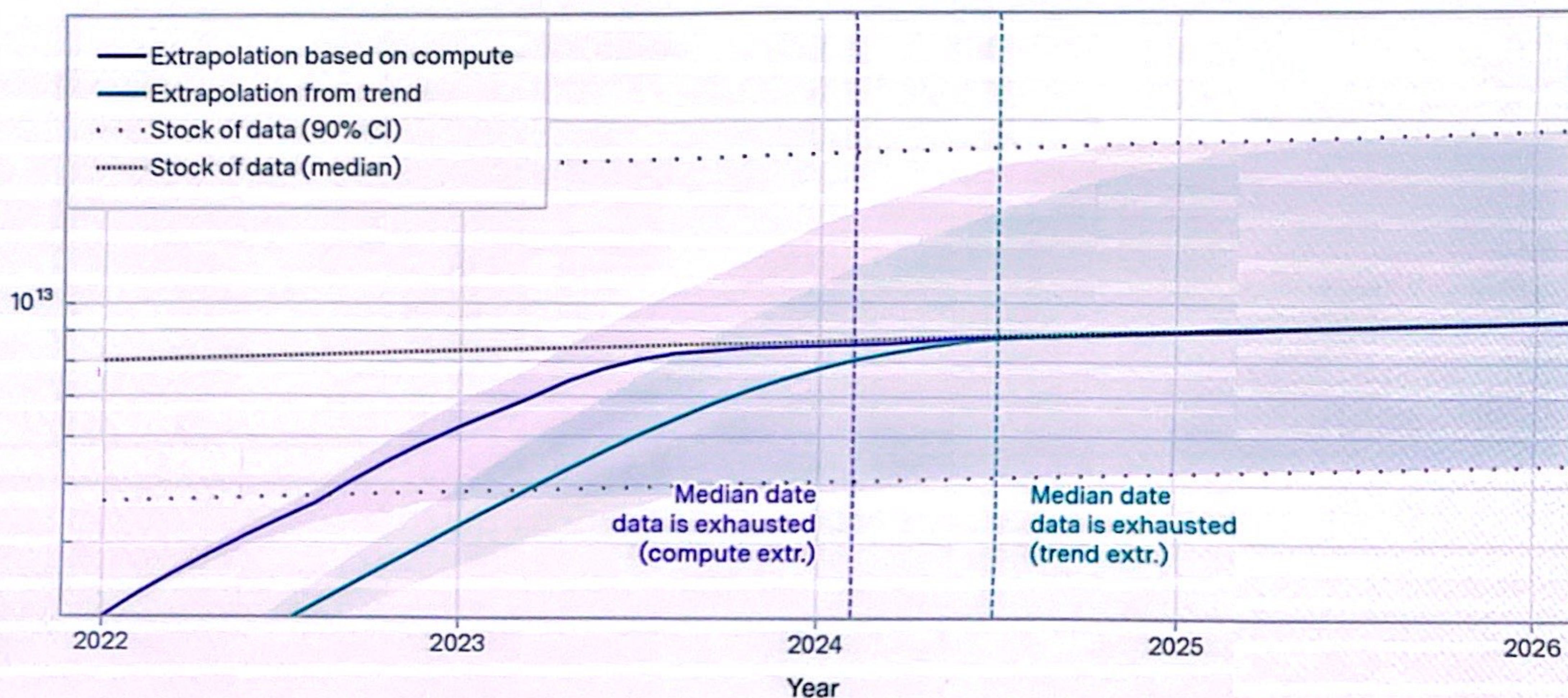
3、大势观澜：人类原生数据即将耗尽，模型技术路径却初见曙光

据人工智能研究机构epoch的研究预测，语言数据可能在2030~2040年耗尽，其中能训练出更优性能的高质量语言数据甚至可能在2026年耗尽。

Projections of data usage (for high-quality language data)

EPOCH AI

Number of words (log)



Scaling Law / 摩尔定律



ChatGPT 3.5

参数1750亿，每个参数用4个Byte表达

$2^{1750 \times 32}$

2^{266}

宇宙原子总量

6、从开源软件商业模式变迁看开源大模型的底层逻辑



MySQL开源策略 (社区版本反馈商业版本)
Terraform by HashiCorp开源调整 (从MPL2.0 到 BSL1.1商业型许可证)

6、从开源软件商业模式变迁看开源大模型的底层逻辑

开源模型项目的许可证调整相对频繁

- 2023年7月份，当时在Huggingface排名首位的Falcon-40B模型（其性能优于LLaMA、MPT、RedPajama 和 StableLM），但是其许可证条款引起比较大的争议。Falcon-40B模型的许可证为其自行起草的“TII Falcon LLM License”。

A) TII Falcon LLM License 要求一旦达到收入门槛，需支付商业用途特许权使用费，而大多数开源许可不需要。TII Falcon LLM License 对如何发布或分发作品也有额外限制，比如要求归属于「Falcon LLM technology from the Technology Innovation Institute」。

B) TII Falcon LLM License 不允许在其他不同的许可下重新授权作品，还明确排除部分了 Falcon LLM 的某些版本 / 大小的许可。

- 2023年9月份，Falcon-180B模型开源出来，该模型采用其自行编制的“FALCON 180B TII LICENSE VERSION 1.0”许可证，该许可证要求商业使用必须另行获得其公司的商业许可。此时，Falcon-40B许可证已经变更为Apache 2.0许可证。

大型模型开源项目与传统开源软件不同，后者可以自组织的进行创作，前者往往都是依托于公司来进行的（计算资源支出），所以开源模型更贴近商业利益。

```
P main • falcon-40b LICENSE.txt
slippyola Add license file 32d2d41 1 day ago
raw history blame contribute diff 10.4 KB

1 TII Falcon LLM License Version 1.0
2 May 2023
3 falconllm.tii.ee
4
5 INTRODUCTORY NOTE
6 This license is, in part, based on the Apache License Version 2.0 (available at http://www.apache.org/licenses/
7
8 TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION
9 1 Definitions.
10 "Commercial Application Addressee" means falconllm.sales@tii.ee.
11 "Commercial Use" means use where there is, or, in relation to a new use case, a reasonable expectation that the
12 "Commercial User" means someone who has applied to make Commercial Use of the Work and been granted permission
13 "Contribution" shall mean any work of authorship, including the original version of the Work and any modification
14 "Contributor" shall mean licensor and any individual or legal entity on behalf of whom a Contribution has been
15 "Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) (i)
16 "Falcon LLM" shall mean only the following releases of TII's Falcon large language models: (i) Falcon-PW-18; (i)
17 "Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by
18 "License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1
19 "Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the lice
20 "Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, inc
21 "Source" form shall mean the preferred form for making modifications, including but not limited to software sou
22 "TII" shall mean the Technology Innovation Institute - Sole Proprietorship L.L.C., or any party nominated in w
23 "Work" shall mean the work of authorship, which in relation to the initial release of Falcon LLM is in Object f
24 "You" (or "Your") shall mean an individual or legal entity exercising permissions granted by this license.
25 2 Grant of Copyright License.
26 2.1 Subject to the terms and conditions of this license, each Contributor hereby grants to You a perpetual, non
27 2.2 Other than where you are a Commercial User in accordance with Section 8, Your copyright license to use the
28 3 Grant of Patent License.
29 3.1 Subject to the terms and conditions of this license, each Contributor hereby grants to You a perpetual, non
30 3.2 Other than where you are a Commercial User in accordance with Section 8, Your patent license to use the Wo
31 4 Redistribution.
32 4.1 Irrespective of whether you are a Commercial User, You may reproduce and distribute copies of the Work or (
33 (a) You must give any other recipients of the Work or Derivative Works a copy of this license; and
34 (b) You must cause any modified files to carry prominent notices stating that You changed the files; and
35 (c) You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, tra
36 (d) If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You d
37 4.2 Subject to Sections 5 and 8 below, where you are a Commercial User, then You may add Your own copyright st
38 4.3 Where You are not a Commercial User, then:
39 (a) You must apply this license in an unmodified form as the terms under which you distribute Your modification
40 (b) Other than in respect of Falcon LLM, which may only be distributed in Object form, where you distribute You
41 (c) You may include a copyright statement recording the nature of your modifications as Derivative Work, but su
42 5 Publication
43 5.1 You shall include prominently in any public statement regarding a Derivative Work the following statement:
44 "[name of relevant Derivative Work] is built using Falcon LLM technology from the Technology Innovation Instit
45 5.2 You may request from TII a reasonably adjusted version of the above statement to suit the publication the
46 5.3 You shall comply with any reasonable request from TII regarding the prominence and phrasing of any referenc
47 6 Submission of Contributions.
48 6.1 Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work i
49 6.2 Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agree
50 7 Trademarks.
51 7.1 Except as required for compliance with Section 5 of this license, this license does not grant permission to
52 8 Commercial Use
53 8.1 Where You wish to make Commercial Use of Falcon LLM or any Derivative Work, You must apply to TII for perm
54 8.2 Where TII grants permission for You to make Commercial Use of the relevant Work, then for that purpose You
55 (a) In its written grant of permission, TII shall set the royalty rate that will apply to you as a Commercial U
56 (b) Each year on the anniversary of the date upon which you were granted permission by TII to make Commercial U
57 (c) Where, on the Anniversary Date, the Attributable Revenue for the preceding 12 months is greater than $1m or
58 9 Disclaimer of Warranty.
59 9.1 Unless required by applicable law or agreed to in writing, licensor provides the Work (and each Contributor
60 10 Limitation of Liability.
61 10.1 In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise,
62 11 Accepting Warranty or Additional Liability.
63 11.1 While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee fo
64
65 END OF TERMS AND CONDITIONS
66
67 APPENDIX: How to apply the TII Falcon LLM license to your work.
68 To apply the TII Falcon LLM license to your work, attach the following boilerplate notice, with the fields encl
69
70 Copyright [yyyy] [name of copyright owner] licensed under the TII Falcon LLM License, Version 1.0 (the "License"
71 You may obtain a copy of the license at falconllm.tii.ee. Unless required by applicable law or agreed to in wri
72 See the license for the specific language governing permissions and limitations under the license.
```

6、从开源软件商业模式变迁看开源大模型的底层逻辑

商业竞争战略决定开源模型许可证条款的演变

对于LLaMA模型，Meta从禁止商业使用转变为允许商业使用（但增加用户数量限制），体现其商业竞争策略，“Android Moment”。“目前 Hugging Face 上流行的10个模型中有 8个是 Llama 2/3的变体”。但是，LLaMA的训练数据及训练代码仍然是闭源的，下游使用者仍然受到Meta的商业制约。

Meta LLaMA LICENSE AGREEMENT

限制

您**不会**、也不会允许、协助或导致任何第三方：

- A. 全部或部分使用、修改、复制、复制、创建软件产品的衍生作品或分发软件产品（或其任何衍生作品、包含软件产品的作品或软件产生的任何数据），以用于 **(i)任何商业或生产目的，(ii) 军事目的或核技术服务，(iii) 监视目的，包括与监视有关的任何研究或开发，(iv) 生物识别处理，(v) 以任何方式侵犯，盗用或以其他方式侵犯任何第三方权利，或 (vi) 以任何方式违反任何适用法律，包括从美国政府禁止的禁运国家/地区访问软件产品，以及违反任何隐私或安全法律、规则、法规、指令；**
- B. 更改或删除软件产品上或内部出现的版权和其他所有权声明；
- C. 利用任何设备、装置、软件或其他方式规避或删除 Meta 使用的与本软件相关的任何安全或保护，或规避或删除任何使用限制，或启用 Meta 禁用的功能；
- D. 对软件产品提供或强加任何改变、限制或与本许可条款不一致的条款。

Meta LLaMA 2 LICENSE AGREEMENT

2.附加商业条款。

如果在 Llama 2 版本发布之日，**被许可方或被许可方的关联公司提供的产品或服务的每月活跃用户数在上一个日历月中超过 7 亿，您必须向以下机构申请许可证：Meta**，Meta 可以自行决定向您授予该权利，并且您无权行使本协议项下的任何权利，除非或直到 Meta 明确授予您此类权利。