



12. September in Erfurt

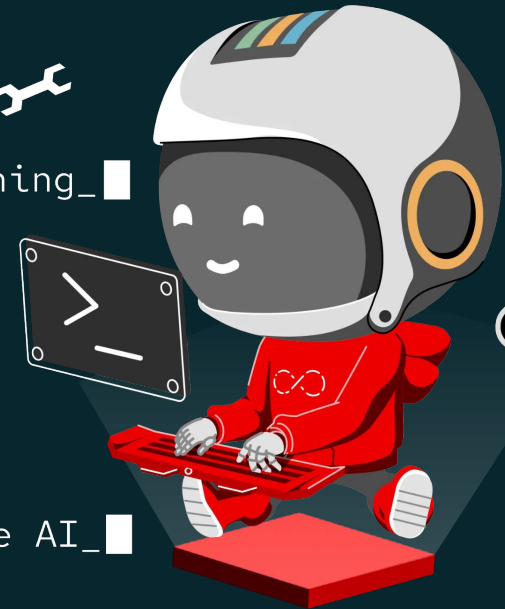
Bitkom Forum Open Source 2024

Die Zukunft von Open Source - fair, geregelt, intelligent

Why you want your AI to be Open Source



Fine-tuning_■



RAG_■



Open-source AI_■

Armin Warda

EMEA FSI Chief Technologist

awarda@redhat.com



Speaker bio

Mr. Armin Warda supports Red Hat's Financial Services customers and partners in the adoption of Red Hat technology, particularly in regards to Operational Efficiency, Security & Compliance as well as on their journey to Hybrid Cloud.

He is currently exploring the impact of European regulations and initiatives on the financial services industry and their IT providers, such as e.g. the Digital Operational Resiliency Act (EU-DORA) and the Artificial Intelligence Act (EU-AIA). Among his areas of interest are also Environment, Social & Governance (ESG) aspects of IT, Trustworthy AI, and the potential transformation of the Payment Industry through the introduction of Central Bank Digital Currencies (CBDCs, such as the Digital-€), as currently being discussed by central banks around the world, the EU and G7.

Prior to his current role he was working for 22 years at Postbank Systems as a Senior IT Architect for Postbank and Deutsche Bank. Armin holds a master degree in Computer Science from the TU Dortmund and also studied at the University College Dublin.



Armin Warda
EMEA FSI Chief Technologist
awarda@redhat.com



Why is NOW a good time for companies to invest in AI?

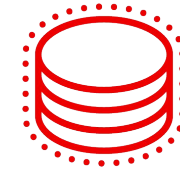
Enterprises are taking the leap



AI technologies are becoming **more accessible and affordable** for businesses of all sizes



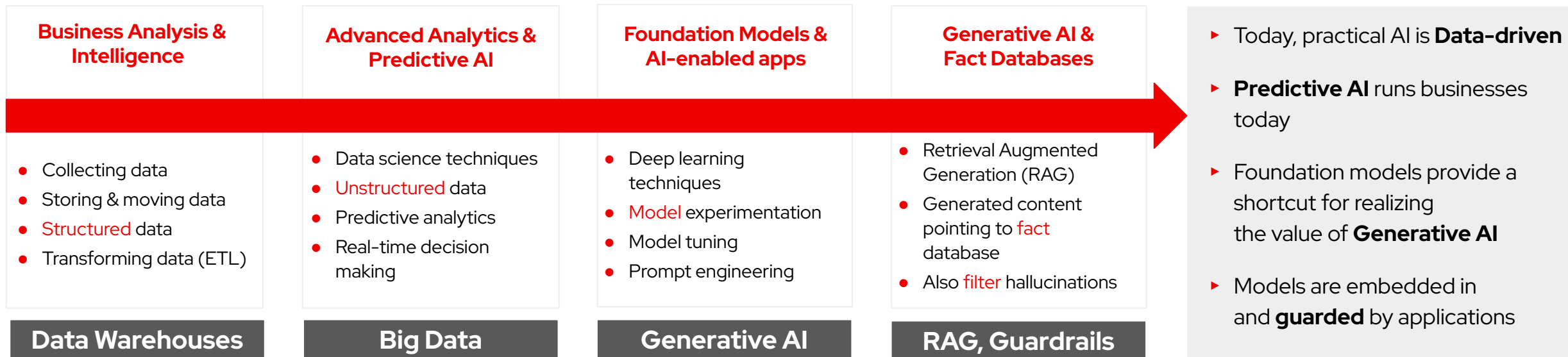
Companies can **realize the value** from AI-enabled applications and AI-support



Organizations are better prepared to manage, transform and **use their ever-increasing data**

AI has undergone significant evolution

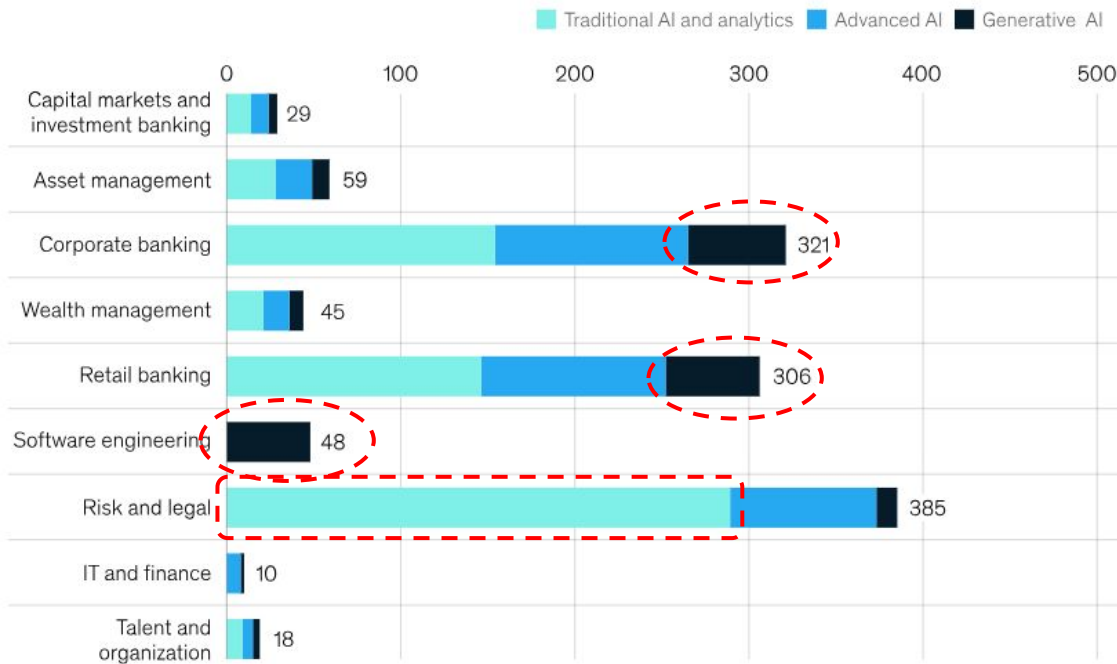
The evolution of **Data-driven** AI: from Business Intelligence to Generative AI



Generative AI in banking has a huge potential for increased productivity

McKinsey Insights: Capturing the full value of generative AI in banking

Value created by AI at stake by segment and function,¹ \$ billion



¹Assumes 0% overlap of traditional AI and generative AI (generative AI assumes the lower end of value at stake), top-down estimation based on projected growth and value pools.
Source: The economic potential of generative AI: The next productivity frontier, McKinsey Global Institute, June 2023; QuantumBlack, AI by McKinsey traditional advanced analytics and AI analysis

Among industry sectors, banking is expected to have one of the largest opportunities, largely from increased productivity

- The economic impact will likely benefit all banking segments and functions, with the greatest absolute gains through Generative AI in the **corporate** and **retail** sectors with \$56 billion and \$54 billion, respectively
- No surprise: software engineering ❤️ LLMs
- **Risk and legal** get largest value from AI, but that's mostly traditional AI

Predictive AI vs. Generative AI

Most common types of AI for business applications

Predictive AI

Predicts or classifies outcomes with models trained on use-case specific data sources



Generative AI

Generates new content with models trained on vast amounts of data from many sources

S I am preparing for a behavioral interview for a marketing role. Can you generate some sample questions about my ability to create and execute successful marketing campaigns? I have experience conducting market research, developing marketing strategies, and analyzing the results of campaigns. Can you also include questions about my creativity and ability to come up with new ideas?"

G Sure! Here are some sample behavioral interview questions about your ability to create and execute successful marketing campaigns:

1. Can you describe a time when you had to develop and implement a marketing campaign from start to finish? What was the campaign for, and what were the results?
2. How do you go about conducting market research when planning a marketing campaign? What tools and techniques do you use to gather information?
3. How do you determine which marketing channels will be most effective for your target audience? Can you give an example of a campaign where you had to make tough choices?
4. How do you measure the success of a marketing campaign? What metrics do you track, and how do you analyze the data to inform future campaigns?
5. Can you describe a time when you had to be creative in order to solve a challenge during a marketing campaign? How did your solution impact the campaign's success?



<https://www.letsdive.io/blog/generative-ai-vs-predictive-ai-all-you-need-to-know#:~:text=Generative%20AI%20is%20primarily%20focused,events%20based%20on%20historical%20data.>

What type of AI drives businesses today?

Predictive AI vs. Generative AI

Most common types of AI for business applications

Predictive AI

Generative AI

What is it for?

Predicts or classifies outcomes with models trained on use-case specific data sources

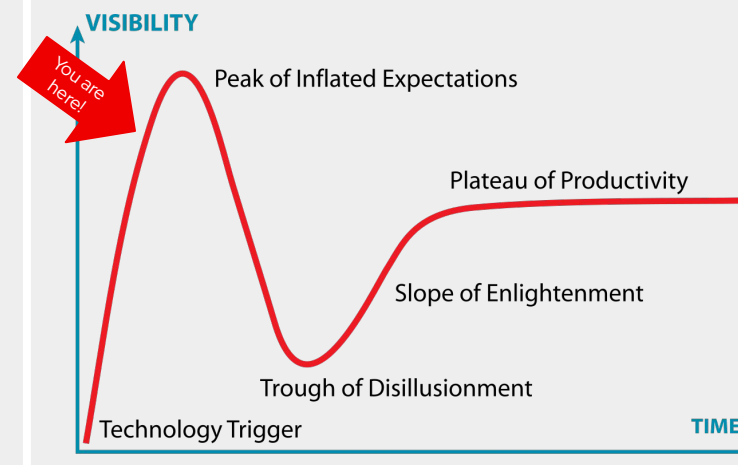
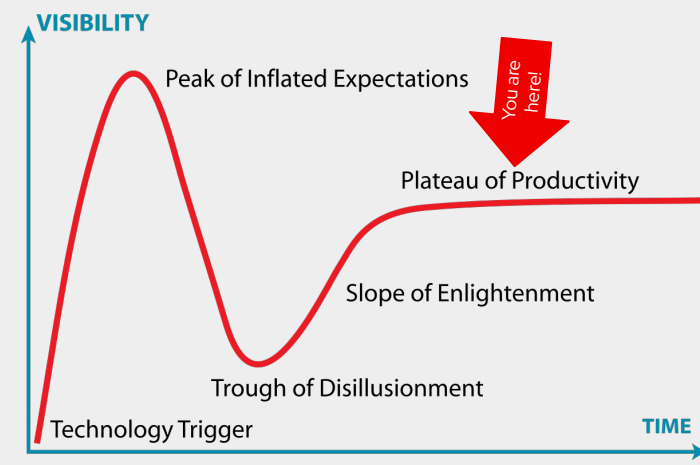
Generates new content with models trained on vast amounts of data from many sources

Penetration

90%

10%

Maturity

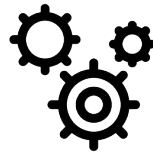


AI Use-Cases in Financial Services



Fraud Management

Anomaly detection,
Countering financial crime
such as money laundering,
terror financing, tax evasion



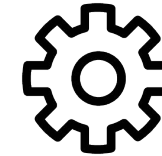
Hyper-Personalization

Improve customer and
employee experience,
Customer Next Best Offer,
Chatbots, Onboarding



Operational Efficiency

Branch Location & Staff
Planning, ATM Cash on
Hand, Call Routing,
Workflow Automation



Risk Analytics

Automated Underwriting
Decisions, Mortgage
Prepayment Analytics,
Credit Scoring



Rage against the machine?

BBC Pandol White

Robotaxis honk at each other



1.00



Crowd burns Waymo

0:07 / 0:41

Crowd burns Waymo self-driving vehicle in San Francisco

CNBC Television
2.64M subscribers

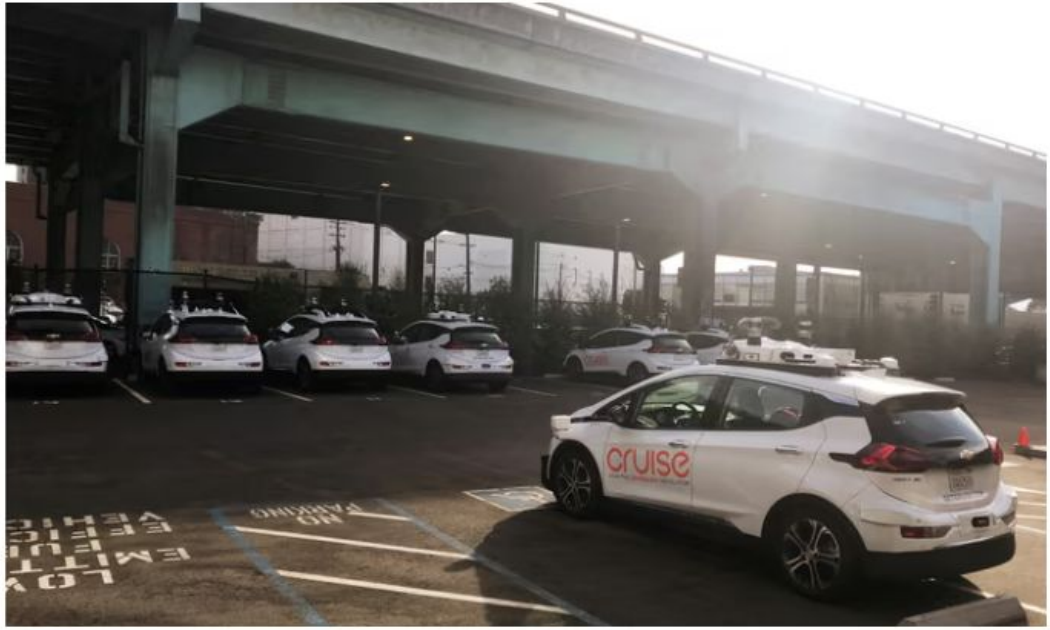
Subscribe

Self-driving cars

This article is more than 3 months old

Cruise recalls all self-driving cars after grisly accident and California ban

All 950 of the General Motors subsidiary's autonomous cars will be taken off roads for a software update



Cruise self-driving cars outside the company's headquarters in San Francisco. Photograph: Heather Somerville/Reuters

Associated Press

Wed 8 Nov 2023 19:17 CET

9

Driverless cars wake residents with nighttime honking

https://www.youtube.com/watch?v=9kn9yu_Cgfc
<https://www.bbc.com/news/videos/c17qqverq99o>
<https://www.theguardian.com/technology/2023/nov/08/cruise-recall-self-driving-cars-gm>



Regulating AI: The EU-AI Act (March 13)



Unregulated, irresponsible or abusive use of AI could lead to negative consequences for individuals or the society, create public opposition and **hinder AI innovation in the EU**.

The EU is committed to strive for a balanced approach to AI

- **Lawful** → accurateness
- **Ethical** → transparency
- **Robust** → fairness
- no (unintended) bias
- security

EU AI Act Requirements:

Explainability, Documentation, Process & Data Governance, Human Oversight, Risk Management, Auditability.

There are some exceptions for AI systems released under **Open Source** licenses.

High Risk
Most regulated AI systems, as these have the potential to cause significant harm if they fail or are misused, e.g. if used in law enforcement or recruiting.

Minimal Risk
All other AI systems, e.g. a spam filter, which can be deployed without additional restrictions.

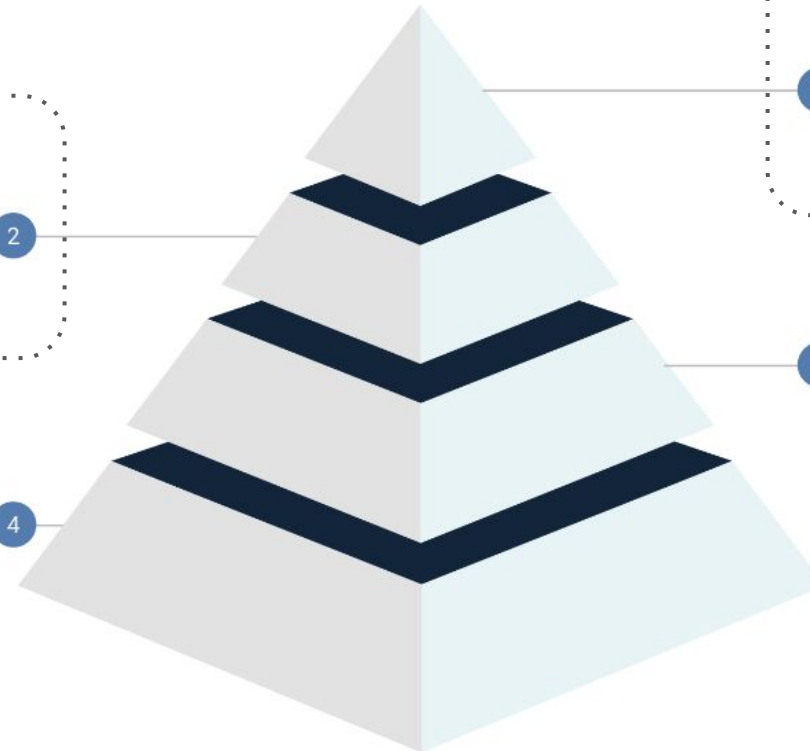
banned:

Unacceptable Risk

1 Highest level of risk prohibited in the EU. Includes AI systems using e.g. subliminal manipulation or general social scoring.

Limited Risk

3 Includes AI systems with a risk of manipulation or deceit, e.g. chatbots or emotion recognition systems. Humans must be informed about their interaction with the AI.



Openwashing?



☰ Openwashing 🗺️ 1 language

Article [Talk](#) Tools

From Wikipedia, the free encyclopedia

Openwashing or open washing (a [compound word](#) modeled on “[whitewash](#)” and derived from “[greenwashing](#)”) is a term to describe presenting something as [open](#), when it is not actually open. In the context of openwashing, ‘open’ refers to transparency, access to information, participation, and knowledge sharing.^[1]

Usage [\[edit\]](#)

The term was coined by Michelle Thorne, an Internet and climate policy scholar,

Rethinking open source generative AI: open-washing and the EU AI Act

Andreas Liesenfeld*
Mark Dingemanse*
andreas.liesenfeld@ru.nl
mark.dingemanse@ru.nl

Centre for Language Studies, Radboud University
Nijmegen, The Netherlands

ABSTRACT

The past year has seen a steep rise in generative AI systems that claim to be open. But how open are they really? The question of what counts as open source in generative AI is poised to take on particular importance in light of the upcoming EU AI Act that regulates open source systems differently, creating an urgent need for practical openness assessment. Here we use an evidence-based framework that distinguishes 14 dimensions of openness, from training datasets to scientific and technical documentation and from licensing to access methods. Surveying over 45 generative AI systems (both text and text-to-image), we find that while the term open source is widely used, many models are ‘open weight’ at best and many providers seek to evade scientific, legal and regulatory scrutiny by withholding information on training and fine-tuning

Janeiro, Brazil. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3630106.3659005>

1 INTRODUCTION

Open generative AI systems are on the rise, with small players and academic initiatives leading the way in open innovation and scientific documentation [20, 32, 61] and several larger corporations joining the fray by releasing models billed as ‘open’. But there are three critical challenges to openness in the domain of generative AI systems. The first is that openness is not a binary feature: today’s transformer-based system architectures and their training procedures are complex, and they can only be classified into open or closed at the price of severe information loss. Secondly, some

Open Source **Software**



Open Source **Hardware**



Open Source **AI/ML Models**



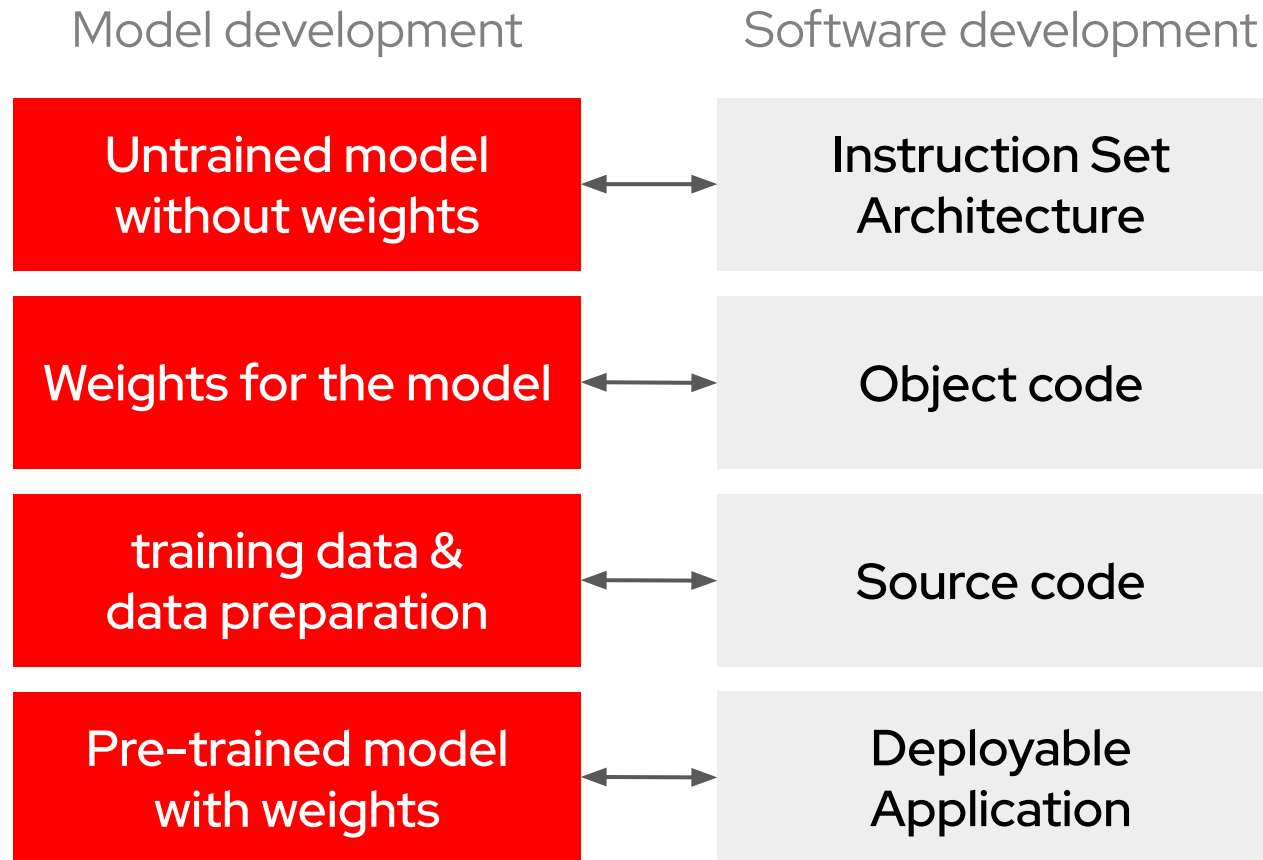
How open are “Open Source” Models?

Open Source Software	Today's “Open Source” LLMs
Frequent releases (sometimes nightly)	Monolithic, irregular releases (e.g. 1y between LLaMA versions)
Incremental contributions	Monolithic development
Feature roadmaps	“Emergent behaviour”, no one knows what’s coming
Community contributions (pull requests)	Largely single-party development (expensive collection of training data)
Contributions from many contributors can be merged and reconciled	Contributions to model, in the form of fine-tuning, are mutually incompatible between contributors, leading to fragmentation in model families
Almost any developer can, in principle, contribute	High barrier to contribution (clusters, GPUs for fine-tuning)

2256	2283	2310	2337	2364	2391	2418	2445	2472	2499	1	28	55	82	109	136	163	190	217	244	271	298	325	1577	1604	1631	1658	1685	1712	1739	1766	1793	1820	1847	1874	1901	
55	2282	2309	2336	2363	2390	2417	2444	2471	2498	1900	27	54	81	108	135	162	189	216	243	270	297	324	326	1603	1630	1657	1684	1711	1738	1765	1792	1819	1846	1873	1275	1277
81	2308	2335	2362	2389	2416	2443	2470	2497	1899	1901	53	80	107	134	161	188	215	242	269	296	323	350	352	1629	1656	1683	1710	1737	1764	1791	1818	1845	1872	1274	1276	1303
07	2334	2361	2388	2415	2442	2469	2496	1898	1925	1927	79	106	133	160	187	214	241	268	295	322	349	351	378	1655	1682	1709	1736	1763	1790	1817	1844	1871	1273	1300	1302	1329
33	2360	2387	2414	2441	2468	2495	1897	1924	1926	1953	105	132	159	186	213	240	267	294	321	348	375	377	404	1681	1708	1735	1762	1789	1816	1843	1870	1272	1299	1301	1328	1355
59	2386	2413	2440	2467	2494	1896	1923	1950	1952	1979	131	158	185	212	239	266	293	320	347	374	376	403	430	1707	1734	1761	1788	1815	1842	1869	1271	1298	1325	1327	1354	1381
85	2412	2439	2466	2493	1895	1922	1949	1951	1978	2005	157	184	211	238	265	292	319	346	373	400	402	429	456	1733	1760	1787	1814	1841	1868	1270	1297	1324	1326	1353	1380	1407
11	2438	2465	2492	1894	1921	1948	1975	1977	2004	2031	183	210	237	264	291	318	345	372	399	401	428	455	482	1759	1786	1813	1840	1867	1269	1296	1323	1350	1352	1379	1406	1433
37	2464	2491	1893	1920	1947	1974	2003	2030	2057	209	236	263	290	317	344	371	398	425	452	479	506	533	560	1785	1812	1839	1866	1268	1295	1322	1349	1351	1378	1405	1432	1459
63	2490	1892	1919	1946	1973	2000	2002	2029	2056	2083	235	262	289	316	343	370	397	424	451	478	505	532	559	1811	1838	1865	1267	1294	1321	1348	1375	1377	1404	1431	1458	1485
89	1891	1918	1945	1972	1999	2001	2028	2055	2082	2109	261	288	315	342	369	396	423	450	477	504	531	558	585	1837	1864	1266	1293	1320	1347	1374	1376	1403	1430	1457	1484	1511
00	1917	1944	1971	1998	2025	2027	2054	2081	2108	2135	287	314	341	368	395	422	449	476	503	530	557	584	1863	1265	1292	1319	1346	1373	1400	1402	1429	1456	1483	1510	1537	
16	1943	1970	1997	2024	2026	2053	2080	2107	2134	2161	2188	340	367	394	421	448	475	502	529	556	583	610	1883	1264	1291	1318	1345	1372	1399	1401	1428	1455	1482	1509	1536	1563
42	1969	1996	2023	2050	2052	2079	2106	2133	2160	2187	339	366	393	420	447	474	501	528	555	582	609	636	1290	1317	1344	1371	1398	1425	1427	1454	1481	1508	1535	1562	1589	
68	1995	2022	2049	2051	2078	2105	2132	2159	2186	2213	365	392	419	446	473	500	527	554	581	608	635	662	1316	1343	1370	1397	1424	1426	1453	1480	1507	1534	1561	1588	1615	
94	2021	2048	2075	2077	2104	2131	2158	2185	2212	2239	391	418	445	472	499	526	553	580	607	634	661	688	1342	1369	1396	1423	1450	1452	1479	1506	1533	1560	1587	1614	1641	
20	2047	2074	2076	2103	2130	2157	2184	2211	2238	2265	417	444	471	498	525	552	579	606	633	660	687	714	1368	1395	1422	1449	1451	1478	1505	1532	1559	1586	1613	1640	1667	
46	2073	2100	2102	2129	2156	2183	2210	2237	2264	2291	443	470	497	524	551	578	605	632	659	686	713	740	1394	1421	1448	1475	1477	1504	1531	1558	1585	1612	1639	1666	1693	
72	2099	2101	2128	2155	2182	2209	2236	2263	2290	2317	469	496	523	550	577	604	631	658	685	712	739	766	1420	1447	1474	1476	1503	1530	1557	1584	1611	1638	1665	1692	1719	
98	2125	2127	2154	2181	2208	2235	2262	2289	2316	2343	495	522	549	576	603	630	657	684	711	738	765	792	1446	1473	1500	1502	1529	1556	1583	1610	1637	1664	1691	1718	1745	
24	2126	2153	2180	2207	2234	2261	2288	2315	2342	2369	521	548	575	602	629	656	683	710	737	764	791	818	1472	1499	1501	1528	1555	1582	1609	1636	1663	1690	1717	1744	1771	
50	2152	2179	2206	2233	2260	2287	2314	2341	2368	2395	547	574	601	628	655	682	709	736	763	790	817	844	1498	1525	1527	1554	1581	1608	1635	1662	1689	1716	1743	1770	1797	
51	2178	2205	2232	2259	2286	2313	2340	2367	2394	2421	573	600	627	654	681	708	735	762	789	816	843	870	1524	1551	1553	1580	1607	1634	1661	1688	1715	1742	1769	1796	1823	
77	2204	2231	2258	2285	2312	2339	2366	2393	2420	2447	601	628	655	682	709	736	763	790	817	844	871	898	1550	1577	1579	1606	1633	1660	1687	1714	1741	1768	1795	1822	1849	
03	2230	2257	2284	2311	2338	2365	2392	2419	2446	2473	629	656	683	710	737	764	791	818	845	872	899	926	1576	1603	1605	1632	1659	1686	1713	1740	1767	1794	1821	1848	1875	
29	381	408	435	462	489	516	543	570	597	624	651	678	705	732	759	786	813	840	867	894	921	948	1602	1629	1631	1658	1685	1712	1739	1766	1793	1820	1847	1874	1901	
55	407	434	461	488	515	542	569	596	623	650	677	704	731	758	785	812	839	866	893	920	947	974	1628	1655	1657	1684	1711	1738	1765	1792	1819	1846	1873	1275	1277	
81	433	460	487	514	541	568	595	622	649	676	703	730	757	784	811	838	865	892	919	946	973	1000	1654	1681	1683	1710	1737	1764	1791	1818	1845	1872	1274	1276	1303	
07	459	486	513	540	567	594	621	648	675	702	729	756	783	810	837	864	891	918	945	972	1000	1027	1680	1707	1709	1736	1763	1790	1817	1844	1871	1273	1300	1302	1329	
33	485	512	539	566	593	620	647	674	701	728	755	782	809	836	863	890	917	944	971	998	1025	1052	1681	1708	1710	1737	1764	1791	1818	1845	1872	1274	1276	1303		
59	511	538	565	592	619	646	673	700	727	754	781	808	835	862	889	916	943	970	997	1024	1051	1078	1682	1709	1711	1738	1765	1792	1819	1846	1873	1275	1277	1304		
85	537	564	591	618	645	672	699	726	753	780	807	834	861	888	915	942	969	996	1023	1050	1077	1104	1683	1710	1712	1739	1766	1793	1820	1847	1874	1901	1928	1955		
11	563	590	617	644	671	698	725	752	779	806	833	860	887	914	941	968	995	1022	1049	1076	1103	1130	1684	1711	1713	1740	1767	1794	1821	1848	1875	1902	1929	1956		
37	589	616	643	670	697	724	751	778	805	832	859	886	913	940	967	994	1021	1048	1075	1102	1129	1156	1685	1712	1714	1741	1768	1795	1822	1849	1876	1903	1930	1957		
63	615	642	669	696	723	750	777	804	831	858	885	912	939	966	993	1020	1047	1074	1101	1128	1155	1182	1686	1713	1715	1742	1769	1796	1823	1850	1877	1904	1931	1958		
89	641	668	695	722	749	776	803	830	857	884	911	938	965	992	1019	1046	1073	1100	1127	1154	1181	1208	1687	1714	1716	1743	1770	1797	1824	1851	1878	1905	1932	1959		
15	667	694	721	748	775	802	829	856	883	910	937	964	991	1018	1045	1072	1099	1126	1153	1180	1207	1234	1688	1715	1717	1744	1771	1798	1825	1852	1879	1906	1933	1960		
41	683	710	737	764	791	818	845	872	899	926	953	980	1007	1034	1061	1088	1115	1142	1169	1196	1223	1250	1689	1716	1718	1745	1772	1799	1826	1853	1880	1907	1934	1961		
67	709	736	763	790	817	844	871	898	925	952	979	1006	1033	1060	1087	1114	1141	1168	1195	1222	1249	1276	1690	1717	1719	1746	1773	1800	1827	1854	1881	1908	1935	1962		
93	735	762	789	816	843	870	897	924	951	978	1005	1032	1059	1086	1113	1140	11																			

What are weights?
Are weights special?

„Weights are code.“



An open source **community** project for GenAI model development

instructlab

Overview Repositories 7 Discussions Projects 1 Packages People 21



InstructLab

Unfollow

README.md

Welcome to the 🐶 InstructLab Project



InstructLab 🐶 uses a novel synthetic data-based alignment tuning method for Large Language Models (LLMs.) The "lab" in InstructLab 🐶 stands for [Large-Scale Alignment for ChatBots](#) [1].

[1] Shivchander Sudalairaj*, Abhishek Bhandwaldar*, Aldo Pareja*, Kai Xu, David D. Cox, Akash Srivastava*. "LAB: Large-Scale Alignment for ChatBots", arXiv preprint arXiv: 2403.01081, 2024. (* denotes equal contributions)

Why InstructLab

There are many projects rapidly embracing and extending permissively licensed AI models, but they are faced with three main challenges:

- Contribution to the models themselves is not possible directly. They show up as forks, which forces consumers to choose a "best-fit" model that isn't easily extensible, and the forks are expensive for model creators to maintain.
- The ability to contribute ideas is limited by a lack of AI/ML expertise. One has to learn how to fork, train, and refine models in order to see their idea move forward. This is a high barrier to entry.
- There is no direct community governance or best practice around review, curation, and distribution of forked models.

Top discussions this past month

Discussions are for sharing announcements, creating conversation in your community, answering questions, and more.

[Start a new discussion](#)

People



[View all](#)

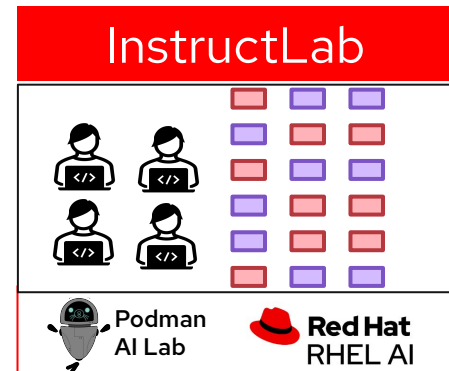
Top languages

- Python
- Shell
- TypeScript
- Jupyter Notebook

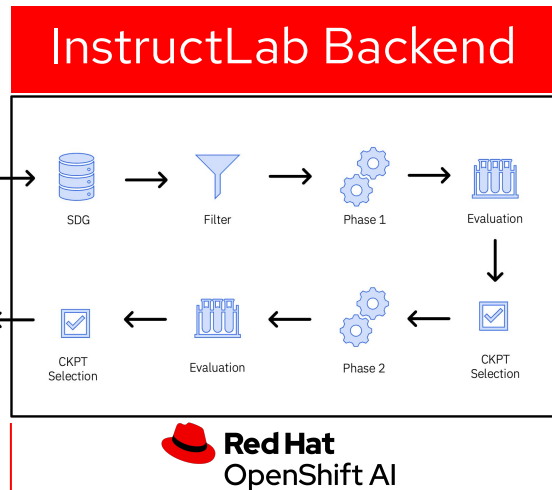
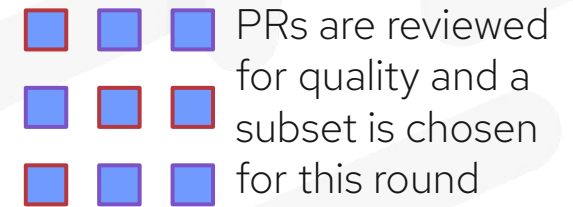
[Report abuse](#)

IBM Research





Triaging Tool & Workflow



Triaged PRs are used to run the backend flow (synthetic data generation + multi-phase training)

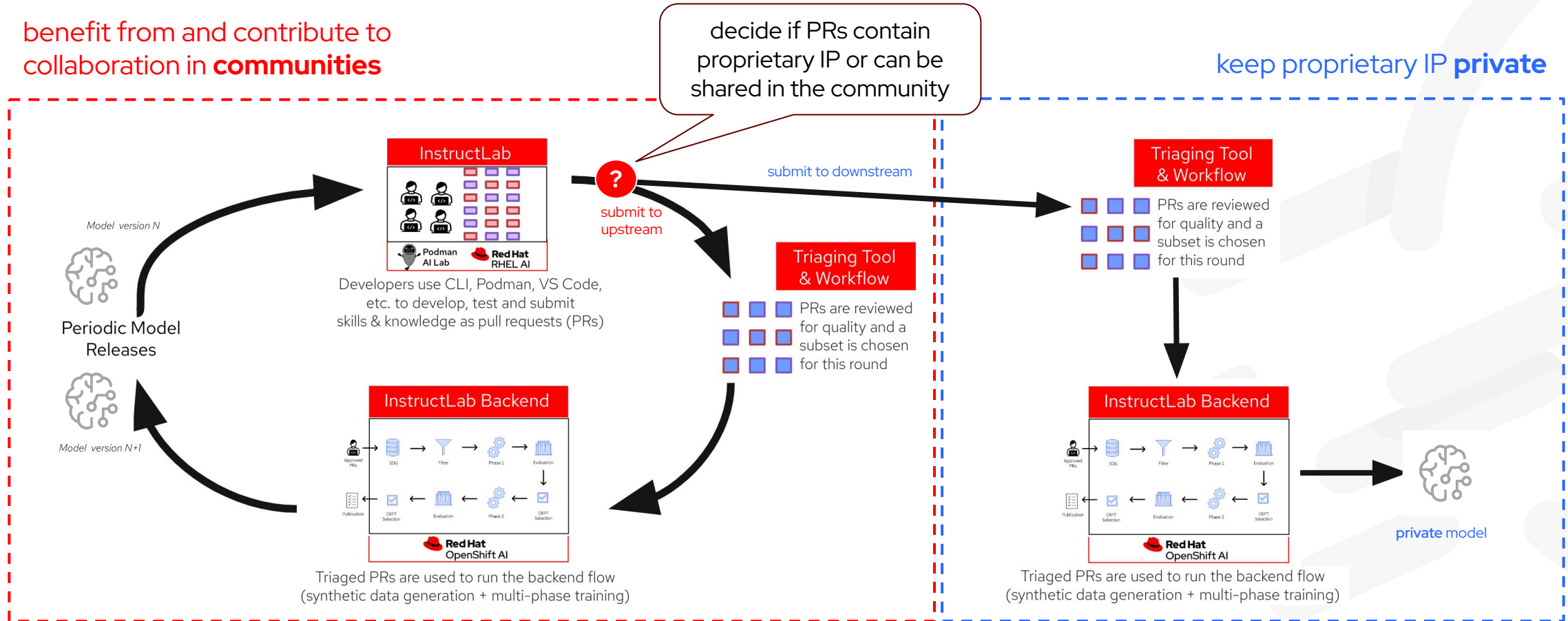
Model version N

Periodic Model Releases

Model version N+1

benefit from and contribute to collaboration in **communities**

keep proprietary IP **private**



Skills and knowledge that can be shared with the community are contributed upstream. These come back for free with the next version of the model, thus reducing the resources required for in-house fine-tuning of the private model, and potentially improved by other collaborators.

Proprietary skills and knowledge, that shall not be shared, are not submitted upstream but retained in-house. These have to be re-added to each new version of the upstream base model.



InstructLab

STEP 1

Learn and experiment via limited desktop-scale training method (qlora) on small datasets.

Podman Desktop integration.

Laptop / desktop



Red Hat Enterprise Linux AI

STEP 2

Production-grade model training using full synthetic data generation, teacher and critic models. Tooling focused on scriptable primitives.

Server / VM



Red Hat OpenShift AI

STEP 3

Production-grade model training as in RHEL AI, using full power of Kubernetes scaling, automation, and MLOps services.

Cluster

Summary



AI/ML benefits businesses

AI-powered intelligent software applications help make better and faster predictions and achieve key business goals.

Open Source AI is winning.



But there are challenges

People, process, compliance and technology challenges can delay deployments. This can have a negative business impact and introduce **new risks**.



Red Hat can help

We have the capabilities and **partnerships** to help speed up AI/ML initiatives, allowing faster delivery of intelligent software applications, and to control the risks.



19. November 2024 Darmstadt

Connect

Thank you



linkedin.com/company/red-hat



facebook.com/redhatinc



youtube.com/user/RedHatVideos



twitter.com/RedHat



awarda@redhat.com

