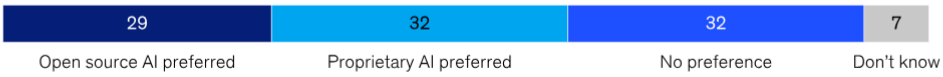
Open Source AI

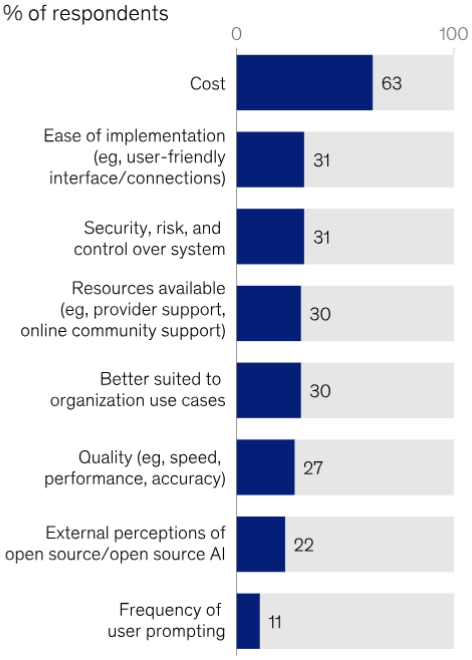
Status Done

Business leaders are intrigued by the possibilities of open source AI, with some reservations.

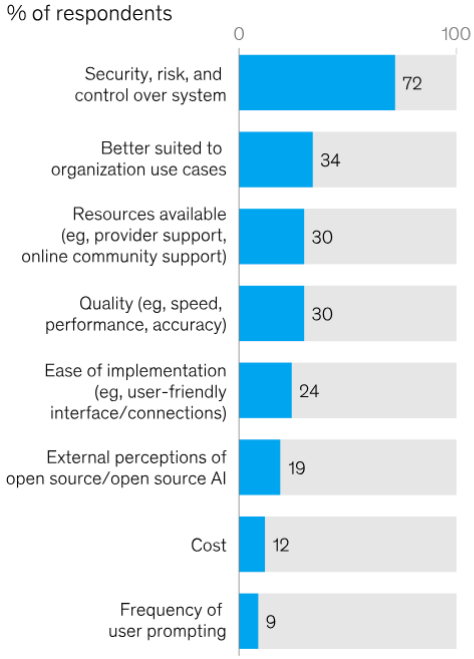
Leadership preferences for open source/proprietary AI,[1] % of respondents



Reasons for preference for open source AI,[2]
% of respondents



Reasons for preference for proprietary AI,[2]
% of respondents



Open source AI: By the numbers

Open source AI use is already widespread

Across several areas of the AI technology stack,

>50%

of respondents reported **using open source AI technologies** (often alongside proprietary AI technologies)

The technology industry leads the way, with

72%

of respondents **using open source AI models**

Organizations that view AI as important to their competitive advantage are >40% more likely to use open source AI models and tools than respondents from other organizations

Developers are particularly excited about open source AI

of developers reported that experience with open source tools is **highly valued in their field**

of respondents reported that developers **more frequently propose using open source AI technologies** than do leaders

Respondents report leaders' preferences are divided

29%

report a preference for open source AI

32%

have no preference

32%

have stated a preference for proprietary AI technologies

Cost perceptions and time to value are driving leadership preferences

60%

of decision makers using open source AI technologies **reported lower implementation costs compared to similar proprietary technologies** and **46% of decision makers see lower maintenance costs**

However, there is still a trade-off in time to value, as

48%

of developers said proprietary AI technologies had **faster time to value**

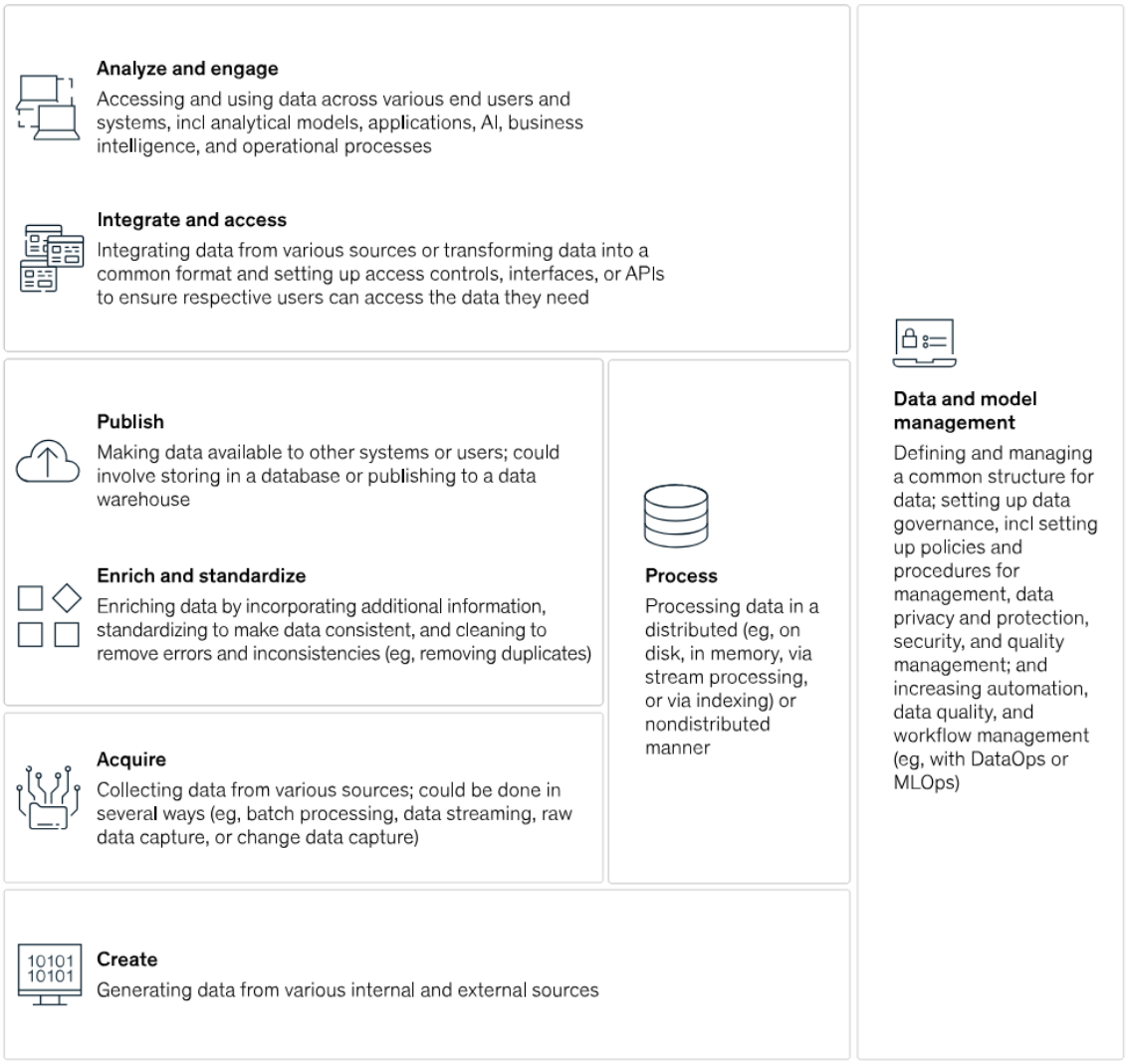
Going forward, open source AI use is likely to increase

76%

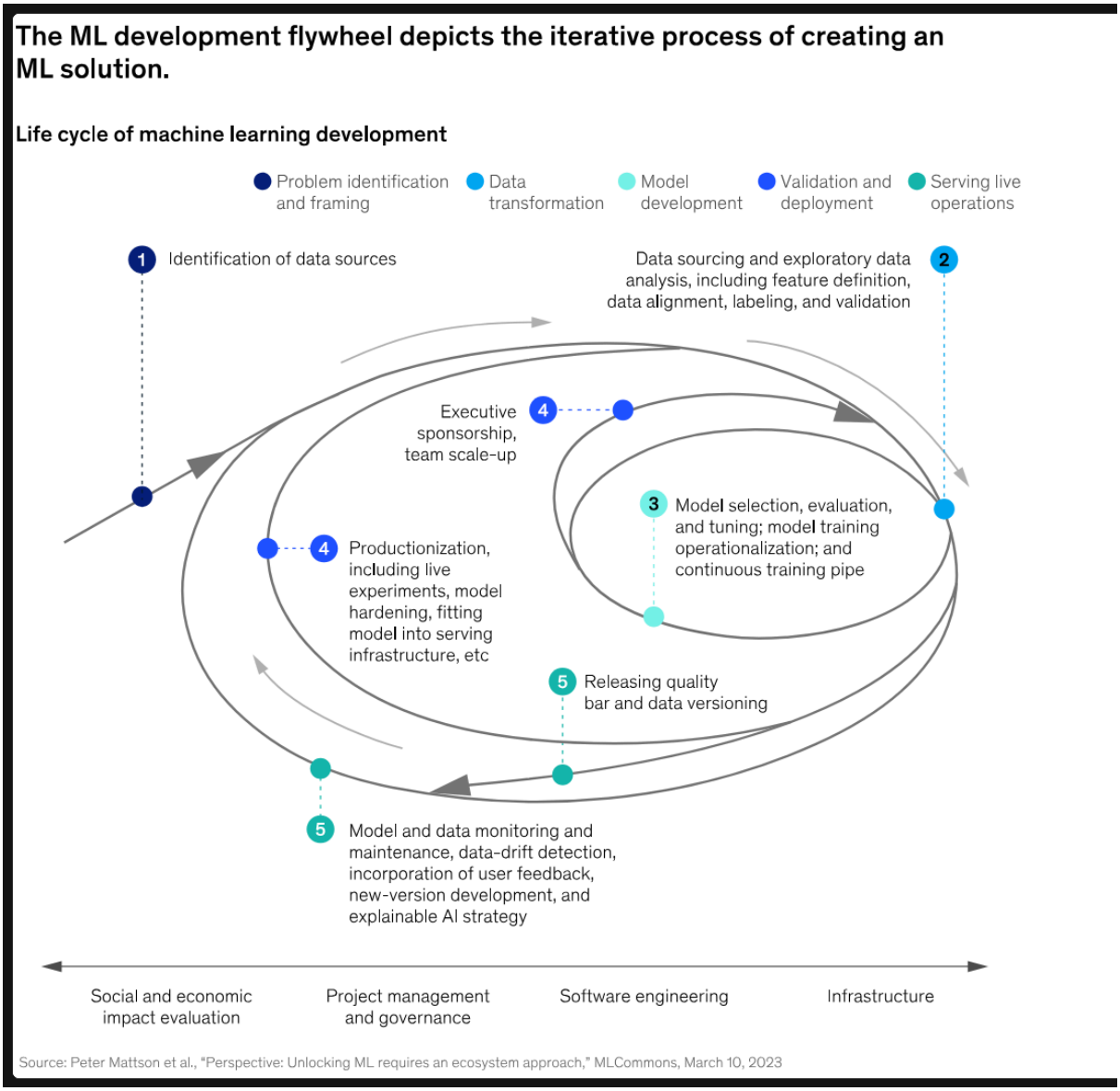
of respondents expect their organization to **increase use of open source AI technologies** over the next several years

McKinsey & Company

A best-in-class reference data and AI architecture provides an essential framework design.



McKinsey & Company



Report titled "The Value of Open Source Software" by Manuel Hoffmann, Frank Nagle, and Yanuo Zhou (Working Paper 24-038), with data backing each claim:

1. Supply-Side Value of Widely-Used OSS

- **Claim**: The cost to recreate all widely-used open source software (OSS) once is estimated at $4.15 billion.
- **Data**: This estimate is derived using the COCOMO II model, which converts lines of code into person-hours and then into dollar values based on global wage data from Salary Expert, weighted by the top 30 countries contributing

to GitHub in 2020 (88% of global activity). The range varies from $1.22 billion (using India's low wages) to $6.22 billion (using U.S. high wages), with $4.15 billion as the global average (Table 2).

2. Demand-Side Value of Widely-Used OSS

- **Claim**: The value of OSS to firms, based on the cost to recreate it for each user if OSS didn't exist, is $8.8 trillion.
- **Data**: This figure scales the supply-side value by usage data from the Census II (2.7 million observations of OSS packages) and BuiltWith (142,000 uses across 3.4 million firm websites), using the COCOMO II model and global wages. The range is $2.59 trillion (India wages) to $13.18 trillion (U.S. wages), with $8.8 trillion as the global average (Table 2).

3. Impact on Firm Software Spending

- **Claim**: Without OSS, firms would need to spend 3.5 times more on software than they currently do.
- **Data**: Current global software spending in 2020 was estimated at $3.4 trillion (derived from Statista's $531.7 billion annual revenue, adjusted to a $1.54 trillion stock over three years, and scaled by U.S. NIPA data showing 45% prepackaged software). Adding the $8.8 trillion demand-side OSS value results in $12.2 trillion total spending, or 3.5 times the current amount (Section 4).

4. Concentration of Value by Programming Language

- **Claim**: The top six programming languages account for 84% of the demand-side value of OSS.
- **Data**: These languages—C, Java, JavaScript, Python, Typescript, and Go—dominate usage in the Census (92% of 2.7 million uses) and BuiltWith (99.97% of 142,000 uses) datasets, contributing 84% of the $8.8 trillion demand-side value (Abstract, Figure 1).

5. Concentration of Value Creation Among Developers

- **Claim**: 96% of the demand-side value is created by just 5% of OSS developers.
- **Data**: Using GHTorrent data (2.3 million commits by 60,000 developers), the study finds that the top 5% (3,000 developers) generate over 96% of the $8.8 trillion demand-side value, with Lorenz curves showing extreme concentration (Figure 4, Section 4).

6. Data Sources

- **Claim**: The study uses two complementary datasets to measure OSS usage comprehensively.
- **Data**:
  - **Census II**: Over 2.7 million observations of OSS packages from software composition analysis firms (Snyk, Synopsys, FOSSA) in 2020, focusing on the top 2,000 packages (70% of total usage).
  - **BuiltWith**: Scans of 8.8 million websites (3.4 million matched to firms via Orbis, Compustat, PitchBook), identifying 142,000 OSS uses, primarily JavaScript libraries (Section 2).

7. Methodology

- **Claim**: A labor market approach estimates OSS value by calculating recreation costs.
- **Data**: The COCOMO II model converts lines of code (261.7 million from Census, 82 million from BuiltWith) into effort (person-hours) and cost (using global wages). Supply-side value assumes one recreation; demand-side scales by unique firm usage (Section 3).

8. Comparison to Previous Studies

- **Claim**: Prior studies overestimated supply-side value by including all OSS, while this study focuses on widely-used OSS and adds demand-side value.
- **Data**: Robbins et al. (2021) and Blind et al. (2021) estimated $78 billion globally for all OSS supply-side value, while this study's $4.15 billion reflects only

widely-used OSS (5.5% of total). No prior demand-side estimate existed due to data limitations (Section 4).

9. Implications

- **Claim**: OSS provides massive economic value, underscoring the need to support its ecosystem, especially key developers.
- **Data**: The $8.8 trillion demand-side value dwarfs current software spending ($3.4 trillion), and 96% of this value hinges on 5% of developers, highlighting dependency on a small group (Sections 4, 5).