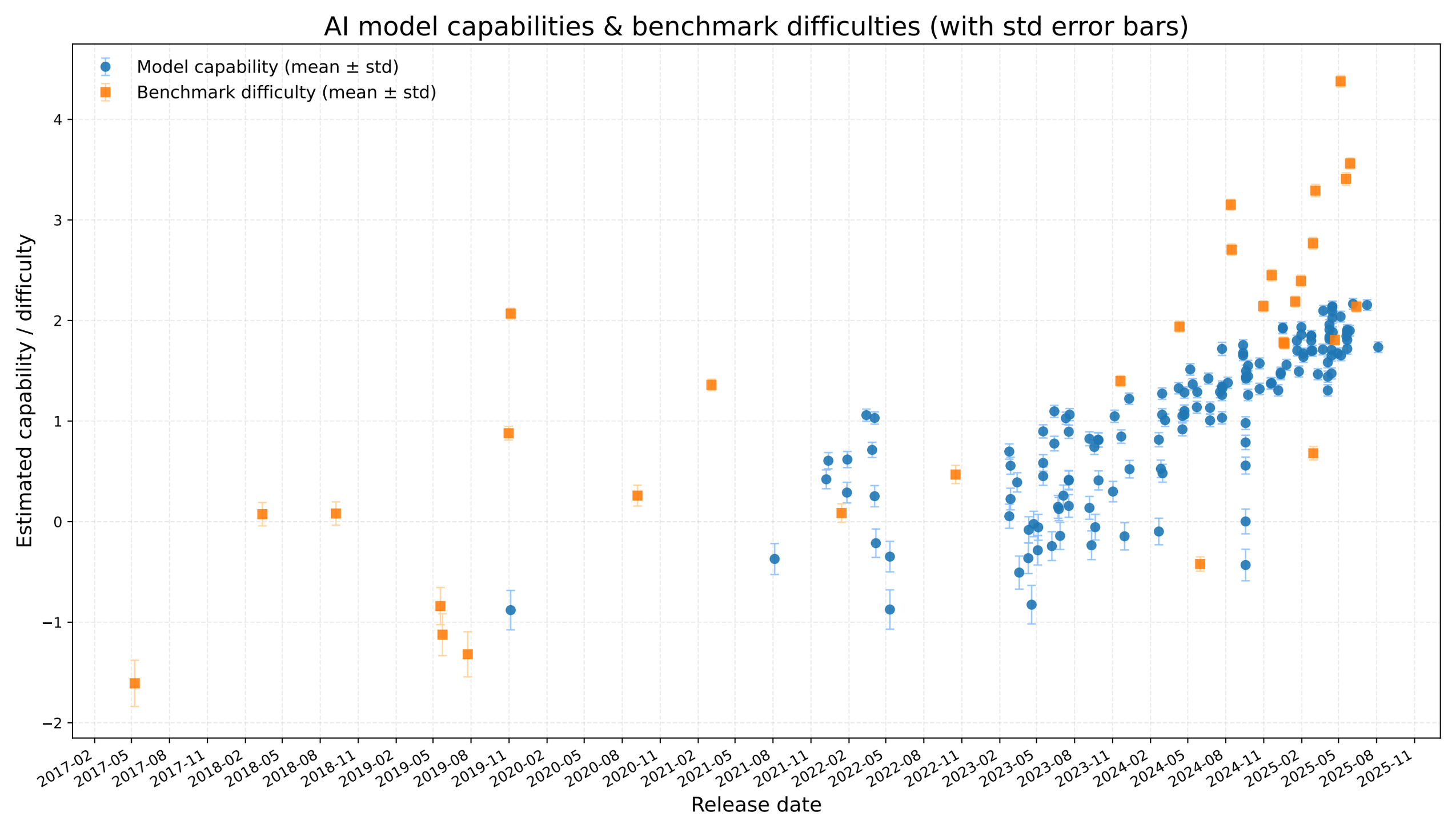
AI model capabilities & benchmark difficulties (with std error bars)
Model capability (mean ± std)
Benchmark difficulty (mean ± std)
Estimated capability / difficulty
Release date
4
3
2
1
0
−1
−2
2017-02
2017-05
2017-08
2017-11
2018-02
2018-05
2018-08
2018-11
2019-02
2019-05
2019-08
2019-11
2020-02
2020-05
2020-08
2020-11
2021-02
2021-05
2021-08
2021-11
2022-02
2022-05
2022-08
2022-11
2023-02
2023-05
2023-08
2023-11
2024-02
2024-05
2024-08
2024-11
2025-02
2025-05
2025-08
2025-11