

Single Day Event

Practical Implementation of AI in Clinical Development: Bridging the Gap from Proof-of-Concept to Production

Hao Chen, Xuejie Zhou

Scientific Programming

Global Statistics and Data Science (GSDS)

BeOne Medicines



Disclaimer

The views and opinions presented in this paper, along with the associated presentation and slides, are solely those of the presenter and do not necessarily represent those of BeOne. This analysis is based entirely on publicly available data, and BeOne does not ensure the accuracy or reliability of the information provided.



Agenda

Background



Challenge & Case Sharing



AI Engineering



Summary



Background - Challenge

■ What We Expect



Tesla Gigafactory 3 Shanghai

<https://www.youtube.com/watch?v=-Ds1xV7M2gI>

- Standardized, accurate, and repeatable
- Well-defined, controlled processes
- Errors easily detected and corrected

■ What We Face



Pony.ai self-drive

<https://www.youtube.com/watch?v=jKFV7qjWaD4>

- Dynamic, ever-changing, and complex
- Human factors and real-world variability
- Rare events and exceptions are critical

The Innovation Boom in China Biotech

Key Takeaways

- Annual revenue from drugs originating in China could jump to an estimated \$34 billion by 2030 and \$220 billion by 2040.
- China is projected to account for 35% of approvals by the U.S. Food and Drug Administration, up from only 5% currently.
- Chinese biotech companies are now increasingly developing novel drugs in areas of high value, including treatments for cancer, diabetes and obesity.
- Global pharmaceutical companies may look for licensing partnerships in China to supplement their drug portfolios.
- Geopolitics remain a key variable to the advance of China's biotech and its integration with the global industry.



Background - Opportunities from AI

Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry and Other Interested Parties

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the Federal Register of this notice increasing the availability of this draft guidance. Submit electronic comments to <https://www.regulations.gov>. Submit written comments to the Center for Drug Evaluation and Research, Food and Drug Administration, 10155 Lauder Lane, 5th, 10th, 12th, 14th, 16th, 18th, 20th, 22nd, 24th, 26th, 28th, 30th, 32nd, 34th, 36th, 38th, 40th, 42nd, 44th, 46th, 48th, 50th, 52nd, 54th, 56th, 58th, 60th, 62nd, 64th, 66th, 68th, 70th, 72nd, 74th, 76th, 78th, 80th, 82nd, 84th, 86th, 88th, 90th, 92nd, 94th, 96th, 98th, 100th, 102nd, 104th, 106th, 108th, 110th, 112th, 114th, 116th, 118th, 120th, 122nd, 124th, 126th, 128th, 130th, 132nd, 134th, 136th, 138th, 140th, 142nd, 144th, 146th, 148th, 150th, 152nd, 154th, 156th, 158th, 160th, 162nd, 164th, 166th, 168th, 170th, 172nd, 174th, 176th, 178th, 180th, 182nd, 184th, 186th, 188th, 190th, 192nd, 194th, 196th, 198th, 200th, 202nd, 204th, 206th, 208th, 210th, 212th, 214th, 216th, 218th, 220th, 222nd, 224th, 226th, 228th, 230th, 232nd, 234th, 236th, 238th, 240th, 242nd, 244th, 246th, 248th, 250th, 252nd, 254th, 256th, 258th, 260th, 262nd, 264th, 266th, 268th, 270th, 272nd, 274th, 276th, 278th, 280th, 282nd, 284th, 286th, 288th, 290th, 292nd, 294th, 296th, 298th, 300th, 302nd, 304th, 306th, 308th, 310th, 312th, 314th, 316th, 318th, 320th, 322nd, 324th, 326th, 328th, 330th, 332nd, 334th, 336th, 338th, 340th, 342nd, 344th, 346th, 348th, 350th, 352nd, 354th, 356th, 358th, 360th, 362nd, 364th, 366th, 368th, 370th, 372nd, 374th, 376th, 378th, 380th, 382nd, 384th, 386th, 388th, 390th, 392nd, 394th, 396th, 398th, 400th, 402nd, 404th, 406th, 408th, 410th, 412th, 414th, 416th, 418th, 420th, 422nd, 424th, 426th, 428th, 430th, 432nd, 434th, 436th, 438th, 440th, 442nd, 444th, 446th, 448th, 450th, 452nd, 454th, 456th, 458th, 460th, 462nd, 464th, 466th, 468th, 470th, 472nd, 474th, 476th, 478th, 480th, 482nd, 484th, 486th, 488th, 490th, 492nd, 494th, 496th, 498th, 500th, 502nd, 504th, 506th, 508th, 510th, 512th, 514th, 516th, 518th, 520th, 522nd, 524th, 526th, 528th, 530th, 532nd, 534th, 536th, 538th, 540th, 542nd, 544th, 546th, 548th, 550th, 552nd, 554th, 556th, 558th, 560th, 562nd, 564th, 566th, 568th, 570th, 572nd, 574th, 576th, 578th, 580th, 582nd, 584th, 586th, 588th, 590th, 592nd, 594th, 596th, 598th, 600th, 602nd, 604th, 606th, 608th, 610th, 612th, 614th, 616th, 618th, 620th, 622nd, 624th, 626th, 628th, 630th, 632nd, 634th, 636th, 638th, 640th, 642nd, 644th, 646th, 648th, 650th, 652nd, 654th, 656th, 658th, 660th, 662nd, 664th, 666th, 668th, 670th, 672nd, 674th, 676th, 678th, 680th, 682nd, 684th, 686th, 688th, 690th, 692nd, 694th, 696th, 698th, 700th, 702nd, 704th, 706th, 708th, 710th, 712th, 714th, 716th, 718th, 720th, 722nd, 724th, 726th, 728th, 730th, 732nd, 734th, 736th, 738th, 740th, 742nd, 744th, 746th, 748th, 750th, 752nd, 754th, 756th, 758th, 760th, 762nd, 764th, 766th, 768th, 770th, 772nd, 774th, 776th, 778th, 780th, 782nd, 784th, 786th, 788th, 790th, 792nd, 794th, 796th, 798th, 800th, 802nd, 804th, 806th, 808th, 810th, 812th, 814th, 816th, 818th, 820th, 822nd, 824th, 826th, 828th, 830th, 832nd, 834th, 836th, 838th, 840th, 842nd, 844th, 846th, 848th, 850th, 852nd, 854th, 856th, 858th, 860th, 862nd, 864th, 866th, 868th, 870th, 872nd, 874th, 876th, 878th, 880th, 882nd, 884th, 886th, 888th, 890th, 892nd, 894th, 896th, 898th, 900th, 902nd, 904th, 906th, 908th, 910th, 912th, 914th, 916th, 918th, 920th, 922nd, 924th, 926th, 928th, 930th, 932nd, 934th, 936th, 938th, 940th, 942nd, 944th, 946th, 948th, 950th, 952nd, 954th, 956th, 958th, 960th, 962nd, 964th, 966th, 968th, 970th, 972nd, 974th, 976th, 978th, 980th, 982nd, 984th, 986th, 988th, 990th, 992nd, 994th, 996th, 998th, 1000th, 1002nd, 1004th, 1006th, 1008th, 1010th, 1012th, 1014th, 1016th, 1018th, 1020th, 1022nd, 1024th, 1026th, 1028th, 1030th, 1032nd, 1034th, 1036th, 1038th, 1040th, 1042nd, 1044th, 1046th, 1048th, 1050th, 1052nd, 1054th, 1056th, 1058th, 1060th, 1062nd, 1064th, 1066th, 1068th, 1070th, 1072nd, 1074th, 1076th, 1078th, 1080th, 1082nd, 1084th, 1086th, 1088th, 1090th, 1092nd, 1094th, 1096th, 1098th, 1100th, 1102nd, 1104th, 1106th, 1108th, 1110th, 1112th, 1114th, 1116th, 1118th, 1120th, 1122nd, 1124th, 1126th, 1128th, 1130th, 1132nd, 1134th, 1136th, 1138th, 1140th, 1142nd, 1144th, 1146th, 1148th, 1150th, 1152nd, 1154th, 1156th, 1158th, 1160th, 1162nd, 1164th, 1166th, 1168th, 1170th, 1172nd, 1174th, 1176th, 1178th, 1180th, 1182nd, 1184th, 1186th, 1188th, 1190th, 1192nd, 1194th, 1196th, 1198th, 1200th, 1202nd, 1204th, 1206th, 1208th, 1210th, 1212th, 1214th, 1216th, 1218th, 1220th, 1222nd, 1224th, 1226th, 1228th, 1230th, 1232nd, 1234th, 1236th, 1238th, 1240th, 1242nd, 1244th, 1246th, 1248th, 1250th, 1252nd, 1254th, 1256th, 1258th, 1260th, 1262nd, 1264th, 1266th, 1268th, 1270th, 1272nd, 1274th, 1276th, 1278th, 1280th, 1282nd, 1284th, 1286th, 1288th, 1290th, 1292nd, 1294th, 1296th, 1298th, 1300th, 1302nd, 1304th, 1306th, 1308th, 1310th, 1312th, 1314th, 1316th, 1318th, 1320th, 1322nd, 1324th, 1326th, 1328th, 1330th, 1332nd, 1334th, 1336th, 1338th, 1340th, 1342nd, 1344th, 1346th, 1348th, 1350th, 1352nd, 1354th, 1356th, 1358th, 1360th, 1362nd, 1364th, 1366th, 1368th, 1370th, 1372nd, 1374th, 1376th, 1378th, 1380th, 1382nd, 1384th, 1386th, 1388th, 1390th, 1392nd, 1394th, 1396th, 1398th, 1400th, 1402nd, 1404th, 1406th, 1408th, 1410th, 1412th, 1414th, 1416th, 1418th, 1420th, 1422nd, 1424th, 1426th, 1428th, 1430th, 1432nd, 1434th, 1436th, 1438th, 1440th, 1442nd, 1444th, 1446th, 1448th, 1450th, 1452nd, 1454th, 1456th, 1458th, 1460th, 1462nd, 1464th, 1466th, 1468th, 1470th, 1472nd, 1474th, 1476th, 1478th, 1480th, 1482nd, 1484th, 1486th, 1488th, 1490th, 1492nd, 1494th, 1496th, 1498th, 1500th, 1502nd, 1504th, 1506th, 1508th, 1510th, 1512th, 1514th, 1516th, 1518th, 1520th, 1522nd, 1524th, 1526th, 1528th, 1530th, 1532nd, 1534th, 1536th, 1538th, 1540th, 1542nd, 1544th, 1546th, 1548th, 1550th, 1552nd, 1554th, 1556th, 1558th, 1560th, 1562nd, 1564th, 1566th, 1568th, 1570th, 1572nd, 1574th, 1576th, 1578th, 1580th, 1582nd, 1584th, 1586th, 1588th, 1590th, 1592nd, 1594th, 1596th, 1598th, 1600th, 1602nd, 1604th, 1606th, 1608th, 1610th, 1612th, 1614th, 1616th, 1618th, 1620th, 1622nd, 1624th, 1626th, 1628th, 1630th, 1632nd, 1634th, 1636th, 1638th, 1640th, 1642nd, 1644th, 1646th, 1648th, 1650th, 1652nd, 1654th, 1656th, 1658th, 1660th, 1662nd, 1664th, 1666th, 1668th, 1670th, 1672nd, 1674th, 1676th, 1678th, 1680th, 1682nd, 1684th, 1686th, 1688th, 1690th, 1692nd, 1694th, 1696th, 1698th, 1700th, 1702nd, 1704th, 1706th, 1708th, 1710th, 1712th, 1714th, 1716th, 1718th, 1720th, 1722nd, 1724th, 1726th, 1728th, 1730th, 1732nd, 1734th, 1736th, 1738th, 1740th, 1742nd, 1744th, 1746th, 1748th, 1750th, 1752nd, 1754th, 1756th, 1758th, 1760th, 1762nd, 1764th, 1766th, 1768th, 1770th, 1772nd, 1774th, 1776th, 1778th, 1780th, 1782nd, 1784th, 1786th, 1788th, 1790th, 1792nd, 1794th, 1796th, 1798th, 1800th, 1802nd, 1804th, 1806th, 1808th, 1810th, 1812th, 1814th, 1816th, 1818th, 1820th, 1822nd, 1824th, 1826th, 1828th, 1830th, 1832nd, 1834th, 1836th, 1838th, 1840th, 1842nd, 1844th, 1846th, 1848th, 1850th, 1852nd, 1854th, 1856th, 1858th, 1860th, 1862nd, 1864th, 1866th, 1868th, 1870th, 1872nd, 1874th, 1876th, 1878th, 1880th, 1882nd, 1884th, 1886th, 1888th, 1890th, 1892nd, 1894th, 1896th, 1898th, 1900th, 1902nd, 1904th, 1906th, 1908th, 1910th, 1912th, 1914th, 1916th, 1918th, 1920th, 1922nd, 1924th, 1926th, 1928th, 1930th, 1932nd, 1934th, 1936th, 1938th, 1940th, 1942nd, 1944th, 1946th, 1948th, 1950th, 1952nd, 1954th, 1956th, 1958th, 1960th, 1962nd, 1964th, 1966th, 1968th, 1970th, 1972nd, 1974th, 1976th, 1978th, 1980th, 1982nd, 1984th, 1986th, 1988th, 1990th, 1992nd, 1994th, 1996th, 1998th, 2000th, 2002nd, 2004th, 2006th, 2008th, 2010th, 2012th, 2014th, 2016th, 2018th, 2020th, 2022nd, 2024th, 2026th, 2028th, 2030th, 2032nd, 2034th, 2036th, 2038th, 2040th, 2042nd, 2044th, 2046th, 2048th, 2050th, 2052nd, 2054th, 2056th, 2058th, 2060th, 2062nd, 2064th, 2066th, 2068th, 2070th, 2072nd, 2074th, 2076th, 2078th, 2080th, 2082nd, 2084th, 2086th, 2088th, 2090th, 2092nd, 2094th, 2096th, 2098th, 2100th, 2102nd, 2104th, 2106th, 2108th, 2110th, 2112th, 2114th, 2116th, 2118th, 2120th, 2122nd, 2124th, 2126th, 2128th, 2130th, 2132nd, 2134th, 2136th, 2138th, 2140th, 2142nd, 2144th, 2146th, 2148th, 2150th, 2152nd, 2154th, 2156th, 2158th, 2160th, 2162nd, 2164th, 2166th, 2168th, 2170th, 2172nd, 2174th, 2176th, 2178th, 2180th, 2182nd, 2184th, 2186th, 2188th, 2190th, 2192nd, 2194th, 2196th, 2198th, 2200th, 2202nd, 2204th, 2206th, 2208th, 2210th, 2212th, 2214th, 2216th, 2218th, 2220th, 2222nd, 2224th, 2226th, 2228th, 2230th, 2232nd, 2234th, 2236th, 2238th, 2240th, 2242nd, 2244th, 2246th, 2248th, 2250th, 2252nd, 2254th, 2256th, 2258th, 2260th, 2262nd, 2264th, 2266th, 2268th, 2270th, 2272nd, 2274th, 2276th, 2278th, 2280th, 2282nd, 2284th, 2286th, 2288th, 2290th, 2292nd, 2294th, 2296th, 2298th, 2300th, 2302nd, 2304th, 2306th, 2308th, 2310th, 2312th, 2314th, 2316th, 2318th, 2320th, 2322nd, 2324th, 2326th, 2328th, 2330th, 2332nd, 2334th, 2336th, 2338th, 2340th, 2342nd, 2344th, 2346th, 2348th, 2350th, 2352nd, 2354th, 2356th, 2358th, 2360th, 2362nd, 2364th, 2366th, 2368th, 2370th, 2372nd, 2374th, 2376th, 2378th, 2380th, 2382nd, 2384th, 2386th, 2388th, 2390th, 2392nd, 2394th, 2396th, 2398th, 2400th, 2402nd, 2404th, 2406th, 2408th, 2410th, 2412th, 2414th, 2416th, 2418th, 2420th, 2422nd, 2424th, 2426th, 2428th, 2430th, 2432nd, 2434th, 2436th, 2438th, 2440th, 2442nd, 2444th, 2446th, 2448th, 2450th, 2452nd, 2454th, 2456th, 2458th, 2460th, 2462nd, 2464th, 2466th, 2468th, 2470th, 2472nd, 2474th, 2476th, 2478th, 2480th, 2482nd, 2484th, 2486th, 2488th, 2490th, 2492nd, 2494th, 2496th, 2498th, 2500th, 2502nd, 2504th, 2506th, 2508th, 2510th, 2512th, 2514th, 2516th, 2518th, 2520th, 2522nd, 2524th, 2526th, 2528th, 2530th, 2532nd, 2534th, 2536th, 2538th, 2540th, 2542nd, 2544th, 2546th, 2548th, 2550th, 2552nd, 2554th, 2556th, 2558th, 2560th, 2562nd, 2564th, 2566th, 2568th, 2570th, 2572nd, 2574th, 2576th, 2578th, 2580th, 2582nd, 2584th, 2586th, 2588th, 2590th, 2592nd, 2594th, 2596th, 2598th, 2600th, 2602nd, 2604th, 2606th, 2608th, 2610th, 2612th, 2614th, 2616th, 2618th, 2620th, 2622nd, 2624th, 2626th, 2628th, 2630th, 2632nd, 2634th, 2636th, 2638th, 2640th, 2642nd, 2644th, 2646th, 2648th, 2650th, 2652nd, 2654th, 2656th, 2658th, 2660th, 2662nd, 2664th, 2666th, 2668th, 2670th, 2672nd, 2674th, 2676th, 2678th, 2680th, 2682nd, 2684th, 2686th, 2688th, 2690th, 2692nd, 2694th, 2696th, 2698th, 2700th, 2702nd, 2704th, 2706th, 2708th, 2710th, 2712th, 2714th, 2716th, 2718th, 2720th, 2722nd, 2724th, 2726th, 2728th, 2730th, 2732nd, 2734th, 2736th, 2738th, 2740th, 2742nd, 2744th, 2746th, 2748th, 2750th, 2752nd, 2754th, 2756th, 2758th, 2760th, 2762nd, 2764th, 2766th, 2768th, 2770th, 2772nd, 2774th, 2776th, 2778th, 2780th, 2782nd, 2784th, 2786th, 2788th, 2790th, 2792nd, 2794th, 2796th, 2798th, 2800th, 2802nd, 2804th, 2806th, 2808th, 2810th, 2812th, 2814th, 2816th, 2818th, 2820th, 2822nd, 2824th, 2826th, 2828th, 2830th, 2832nd, 2834th, 2836th, 2838th, 2840th, 2842nd, 2844th, 2846th, 2848th, 2850th, 2852nd, 2854th, 2856th, 2858th, 2860th, 2862nd, 2864th, 2866th, 2868th, 2870th, 2872nd, 2874th, 2876th, 2878th, 2880th, 2882nd, 2884th, 2886th, 2888th, 2890th, 2892nd, 2894th, 2896th, 2898th, 2900th, 2902nd, 2904th, 2906th, 2908th, 2910th, 2912th, 2914th, 2916th, 2918th, 2920th, 2922nd, 2924th, 2926th, 2928th, 2930th, 2932nd, 2934th, 2936th, 2938th, 2940th, 2942nd, 2944th, 2946th, 2948th, 2950th, 2952nd, 2954th, 2956th, 2958th, 2960th, 2962nd, 2964th, 2966th, 2968th, 2970th, 2972nd, 2974th, 2976th, 2978th, 2980th, 2982nd, 2984th, 2986th, 2988th, 2990th, 2992nd, 2994th, 2996th, 2998th, 3000th, 3002nd, 3004th, 3006th, 3008th, 3010th, 3012th, 3014th, 3016th, 3018th, 3020th, 3022nd, 3024th, 3026th, 3028th, 3030th, 3032nd, 3034th, 3036th, 3038th, 3040th, 3042nd, 3044th, 3046th, 3048th, 3050th, 3052nd, 3054th, 3056th, 3058th, 3060th, 3062nd, 3064th, 3066th, 3068th, 3070th, 3072nd, 3074th, 3076th, 3078th, 3080th, 3082nd, 3084th, 3086th, 3088th, 3090th, 3092nd, 3094th, 3096th, 3098th, 3100th, 3102nd, 3104th, 3106th, 3108th, 3110th, 3112th, 3114th, 3116th, 3118th, 3120th, 3122nd, 3124th, 3126th, 3128th, 3130th, 3132nd, 3134th, 3136th, 3138th, 3140th, 3142nd, 3144th, 3146th, 3148th, 3150th, 3152nd, 3154th, 3156th, 3158th, 3160th, 3162nd, 3164th, 3166th, 3168th, 3170th, 3172nd, 3174th, 3176th, 3178th, 3180th, 3182nd, 3184th, 3186th, 3188th, 3190th, 3192nd, 3194th, 3196th, 3198th, 3200th, 3202nd, 3204th, 3206th, 3208th, 3210th, 3212th, 3214th, 3216th, 3218th, 3220th, 3222nd, 3224th, 3226th, 3228th, 3230th, 3232nd, 3234th, 3236th, 3238th, 3240th, 3242nd, 3244th, 3246th, 3248th, 3250th, 3252nd, 3254th, 3256th, 3258th, 3260th, 3262nd, 3264th, 3266th, 3268th, 3270th, 3272nd, 3274th, 3276th, 3278th, 3280th, 3282nd, 3284th, 3286th, 3288th, 3290th, 3292nd, 3294th, 3296th, 3298th, 3300th, 3302nd, 3304th, 3306th, 3308th, 3310th, 3312th, 3314th, 3316th, 3318th, 3320th, 3322nd, 3324th, 3326th, 3328th, 3330th, 3332nd, 3334th, 3336th, 3338th, 3340th, 3342nd, 3344th, 3346th, 3348th, 3350th, 3352nd, 3354th, 3356th, 3358th, 3360th, 3362nd, 3364th, 3366th, 3368th, 3370th, 3372nd, 3374th, 3376th, 3378th, 3380th, 3382nd, 3384th, 3386th, 3388th, 3390th, 3392nd, 3394th, 3396th, 3398th, 3400th, 3402nd, 3404th, 3406th, 3408th, 3410th, 3412th, 3414th, 3416th, 3418th, 3420th, 3422nd, 3424th, 3426th, 3428th, 3430th, 3432nd, 3434th, 3436th, 3438th, 3440th, 3442nd, 3444th, 3446th, 3448th, 3450th, 3452nd, 3454th, 3456th, 3458th, 3460th, 3462nd, 3464th, 3466th, 3468th, 3470th, 3472nd, 3474th, 3476th, 3478th, 3480th, 3482nd, 3484th, 3486th, 3488th, 3490th, 3492nd, 3494th, 3496th, 3498th, 3500th, 3502nd, 3504th, 3506th, 3508th, 3510th, 3512th, 3514th, 3516th, 3518th, 3520th, 3522nd, 3524th, 3526th, 3528th, 3530th, 3532nd, 3534th, 3536th, 3538th, 3540th, 3542nd, 3544th, 3546th, 3548th, 3550th, 3552nd, 3554th, 3556th, 3558th, 3560th, 3562nd, 3564th, 3566th, 3568th, 3570th, 3572nd, 3574th, 3576th, 3578th, 3580th, 3582nd, 3584th, 3586th, 3588th, 3590th, 3592nd, 3594th, 3596th, 3598th, 3600th, 3602nd, 3604th, 3606th



Background - AI Implementation from POC to Production

MIT

The GenAI Divide STATE OF AI IN BUSINESS 2025

3.2 THE PILOT-TO-PRODUCTION CHASM

Takeaway: The GenAI Divide is starkest in deployment rates, only 5% of custom enterprise AI tools reach production. Chatbots succeed because they're easy to try and flexible, but fail in critical workflows due to lack of memory and customization. This fundamental gap explains why most organizations remain on the wrong side of the divide.

Our research reveals a steep drop-off between investigations of GenAI adoption tools and pilots and actual implementations, with significant variation between generic and custom solutions.

Research Limitations: These figures are directionally accurate based on individual interviews rather than official company reporting. Sample sizes vary by category, and success definitions may differ across organizations.

Exhibit: The steep drop from pilots to production for task-specific GenAI tools reveals the GenAI divide

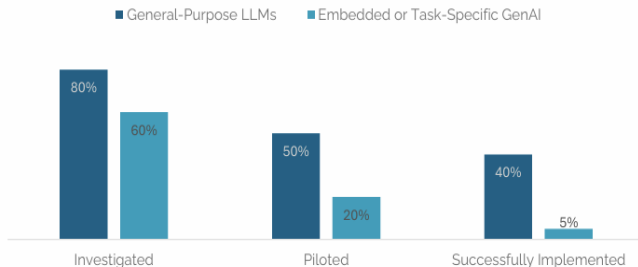


Exhibit: Why GenAI pilots fail: top barriers to scaling AI in the enterprise

Users were asked to rate each issue on a scale of 1-10



MIT report: 95% of generative AI pilots at companies are failing



Background - AI Implementation Strategy

Vision -> Insight -> Execution

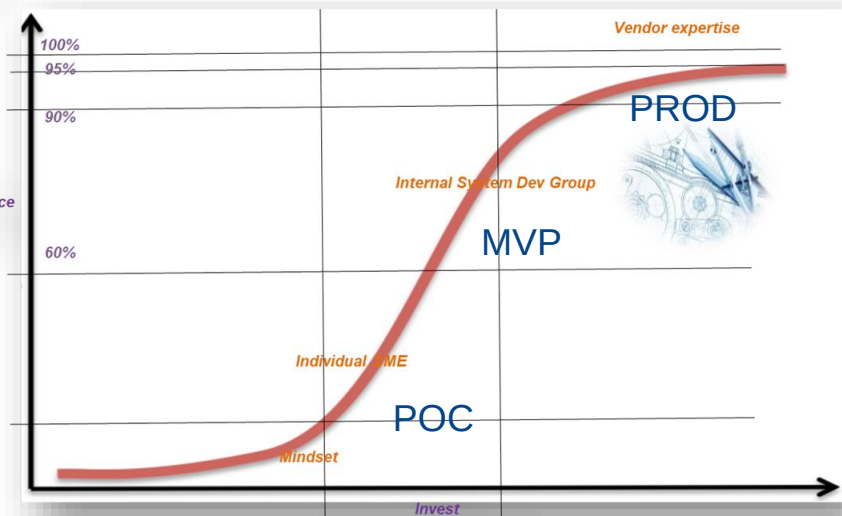
Tech Feasibility

Business Value



Clear Requirements
Good Training Data

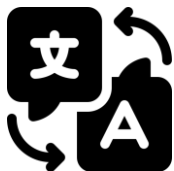
Performance





Background - Our Experience

Translation ★★ ★



AI-enhanced multilingual translation platform for clinical documents

CodeGenius ★★ ★



AI-powered IDE for automated SAS programming and optimization

MedCoding ★★



Automated medical coding platform with clustering and scoring capabilities

SDTM DE ★



AI-driven R-based platform for automated SDTM dataset generation

- Platform Deployment: 2 PROD, 1 UAT, 1 DEV
- Target Users: Medical Writing, Programing, Data Management
- Programming Languages: R, Python, C#
- ML/AI: Semantic Search, Embedding, Clustering, Segmentation, RAG, Prompt Engineering, Agent, MCP, Fine-Tuning
- Integration Capabilities: MS Azure, Databricks, ElasticSearch, Posit
- LLMs: Azure OpenAI, Claude, Deepseek, Qwen



Background - Comparative Pros & Cons of LLM Implementation Methodologies

Method	Pros	Cons	MedCoding	Translation	Code Genius	SDTM DE
Prompt	- Instant deployment with no training overhead - Highly flexible and easy to iterate - Minimal compute cost	- Limited control over output consistency - Can require complex engineering for nuance - Prone to hallucinations	x	x	x	x
RAG (Retrieval-Augmented Generation)	- Grounded responses with up-to-date, factual information - Reduces hallucinations without retraining - Scalable to large corpora	- Complex setup (indexing, embeddings, retrieval) - Additional latency at query time - Maintenance of data pipeline required	x	x	x	x
Agent	- Automates multi-step workflows and integrates external tools - Modular combination of strategies (prompt, RAG, APIs) - Facilitates orchestration of complex tasks	- High development complexity - Requires monitoring and error handling - Potentially brittle if downstream APIs change	x	x		x
Fine-Tuning	- Deep customization for domain-specific tasks - Consistent, high-quality outputs - Reduced reliance on prompt engineering for niche contexts	- Significant compute and data requirements - Long retraining cycles - Less flexible—model updates require re-training		x		x



Challenge & Case Sharing



Phase	Challenge	Translation	CodeGenius	MedCoding	SDTM
Planning & Design	Real-world business value or ROI		✓		
	Success metrics		✓		
	Interaction/UX design	✓		✓	
POC & Validation	Models vs. whole system engineering			✓	
	Depth-first approach	✓			✓
Development	Integration with legacy systems	✓		✓	
	Systematic context engineering	✓		✓	✓
Production Deployment	Security and compliance risks	✓		✓	✓
	Change management and employee buy-in		✓		✓
Iteration & Maintenance	Expectation management			✓	
	Model performance	✓	✓	✓	✓
	Competitive moat	✓	✓	✓	✓

Planning & Design - Interaction/UX design

Translation API

Translation Add-in

Show Difference for review AI results easily

MedCoding

Customized Prompt & Clustering & Scoring

Provide self-defined prompt, order and flag results for review conveniently

Site Number	Subject	ICF page name	Record Position	Record Change	CRF Items (Exclusions)	PT CODE	HLT	HLT CODE	HLST	CODE	SOC	SOC CODE	Version	Medical History Comment	Coder Comment	At Coding Method 1	At Coding 1	At Coding Method 2	At Coding 2	Condition Name	Condition Score	Condition Rating
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
VOM6796	Vomiting	10047700	Vomiting	10047700	Neonates and infants symptoms	10028817	symptoms	10018612	disorders	10017947	27			DOM	Vomiting	Vomiting	At-Syn	Reference	Vomiting	2-6.1	100	Excellent
VOM6796	Vomiting	10047700	Vomiting	10047700	Neonates and infants symptoms	10028817	symptoms	10018612	disorders	10017947	27			DOM	Vomiting	Vomiting	At-Syn	Reference	Vomiting	2-6.1	100	Excellent
VOM6796	Vomiting	10047700	Vomiting	10047700	Neonates and infants symptoms	10028817	symptoms	10018612	disorders	10017947	27			DOM	Vomiting	Vomiting	At-Syn	Reference	Vomiting	2-6.1	100	Excellent
VOM6796	Vomiting	10047700	Vomiting	10047700	Neonates and infants symptoms	10028817	symptoms	10018612	disorders	10017947	27			DOM	Vomiting	Vomiting	At-Syn	Reference	Vomiting	2-6.1	100	Excellent
NEUTROPHILIC COLIC DECREASED	Neutrophilic colic decreased	10029394	decreased	10029394	White blood cell analysis	10018612	analysis	10018612	disorders	10018612	disorders	10018612	27	DOM	Neutrophilic colic decreased	Neutrophilic colic decreased	At-Syn	Neutrophilic colic decreased	Neutrophilic colic decreased	2-6.1	100	Excellent

Medcoding Report Example

Source Text:

TEAEs leading to dose modification of tislelizumab were common in both parts (Table 14.3.1.2.6.1).

Translation Memory (English):

TEAE Leading to Dose Modification of Tislelizumab

English Differences (Memory – Source):

TEAE TEAEs Leading to Dose Modification of Tislelizumab **were common in both parts (Table 14.3.1.2.6.1)**

All Translation:

导致替雷利单抗剂量调整的TEAE在两个部分中均常见 (表14.3.1.2.6.1)。

Translation Memory (Chinese):

导致替雷利单抗的给药调整

English Differences (Memory – Translation):

导致替雷利单抗剂量调整的TEAE在两个部分中均常见 (表14.3.1.2.6.1)。



Challenge & Case Sharing

POC & Validation - Depth-first approach

- Start with ONE complete, production-ready solution
 - eCRT Translation Platform -> R&D Translation Platform
 - Focus deeply on single use case: Protocol translation
 - Validate thoroughly before expansion
- Integrate Knowledge base into the platform
 - Leverage domain specific knowledge (e.g., CDISC SDTM MCP)



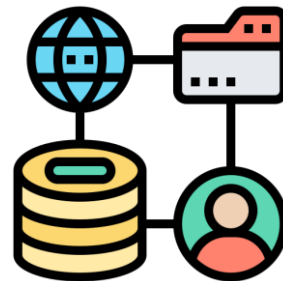


Challenge & Case Sharing

Development - Integration complexity with legacy systems

Context:

- Trados: Medical writing translation software  Trados Studio
- Our asset: Python-based translation function
- Gap: Trados C# plugin framework



Solution Strategy:

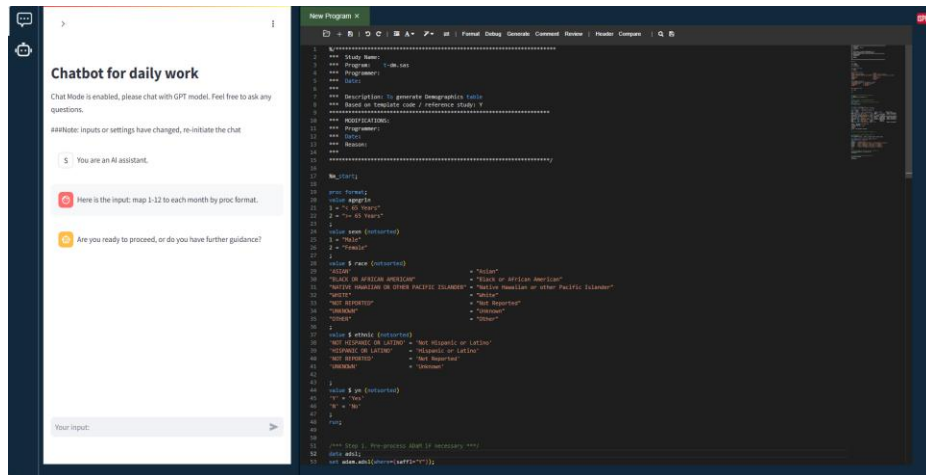


- Wrapped Python function as API on Databricks **databricks**
- Engaged Trados technical support, rapid framework learning through vendor guidance
- Built C# plugin with new features based on user requirements
 - Translation Memory Upload/Multi-selection/priority list
 - Difference between source and matched source & target and matched target



Challenge & Case Sharing

Production Deployment - Change management and employee buy-in



CodeGenius UI

User Adoption Barriers:

- Habit disruption: SAS deeply ingrained in daily work
- Missing feature: Cannot execute code directly (critical gap)

Lack of Change Management:

- Insufficient feedback collection at release
- Limited tracking of adoption metrics

Platform Enhancement:

- Embedded company standard macros, template code
- Iterative improvements addressing user pain points
- GxP Validation for SAS engine connection

Platform Trace:

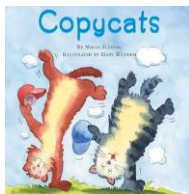
- User Champion Team
- AI API use metrics



Challenge & Case Sharing

Iteration & Maintenance - Competitive moat

Feature
Copycats



Tech
Giant



Legacy
Players



Knowledge

- Translation Memory
- Term Base
- Historical MedCoding Reports
- Study & Standard SAS Template Code
- ...

Algorithm

- Domain-specific model fine-tuning (e.g., translation model)
- Customized AI pipelines (SDTM / MedCoding)
- ...

Ecosystem Integration

- Deep integrations with industry platforms + End-to-end
- Domain-specific API services
- ...

Expert Team

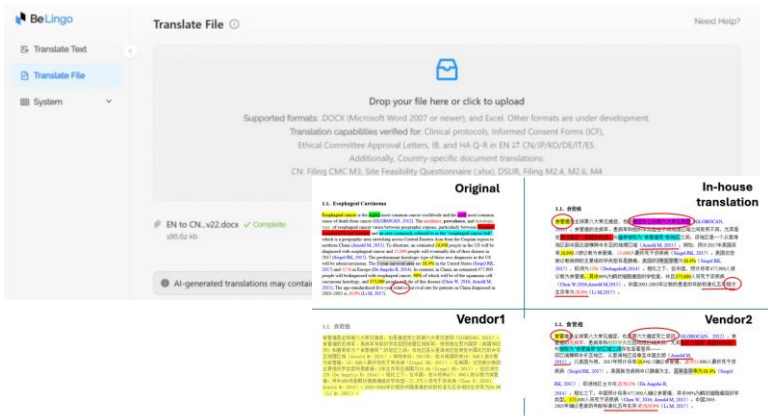
- Multidisciplinary expert team covering the full clinical trial lifecycle (e.g., medical translator, coders, biostatistician, statistical programmers, CDISC specialists)
- ...



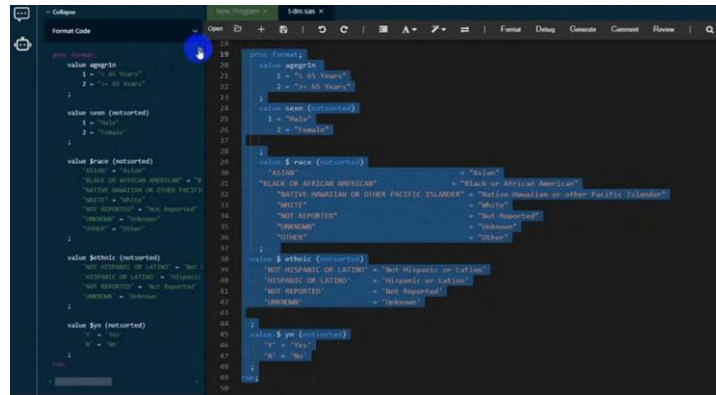
Challenge & Case Sharing



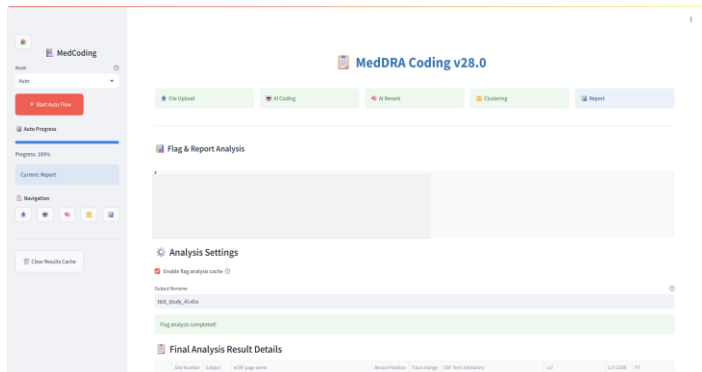
Translation



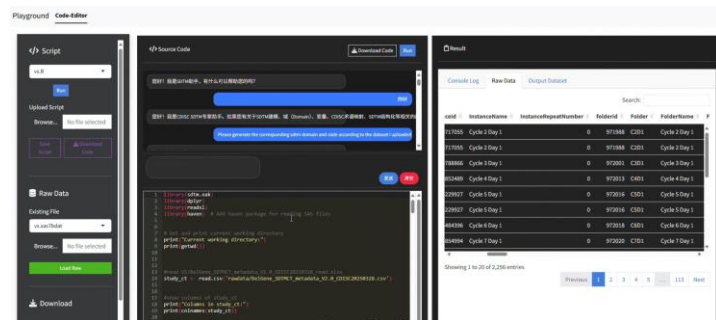
CodeGenius
SAS PROGRAMMING



MedCoding



SDTM Digital Employee



Design Demo



AI Engineering - Data Processing

Transform noisy raw data into searchable and graded data.

Data Clean: Through pre-defined rules and machine learning models, automatically clean and verify data to grade data quality.

Data Strategy: For different quality levels of data, different training strategies are used to maximize the value of data.

Rule-Based vs Gold-Trained vs LLM



Business Rules

Utilizes **deterministic regex patterns** to match and correct common misspellings and known artefacts.



Gold-Trained Model

A domain-specific model fine-tuned exclusively on **high-quality (gold) data**. Data cleaning and grading based on probability distribution differences



General LLM

Leverages a large language model with **in-context learning** to denoise entire sentences. It generalizes well to unseen corruptions.



AI Engineering - Model Training: Split-Strategy Training



Gold Data

High-quality, manually labelled data used for **Supervised Fine-Tuning (SFT)**.

Method:

Standard Cross-Entropy Loss



Noisy Data

Silver data with label noise. Optimized with Robust Loss or RL.

Method:

Robust Loss or RL



Unlabeled Text

Raw text for **Continued Pre-Training (CPT)** to broaden context.

Method:

Masked-Language or Next Token Predict Objective



AI Engineering - Model Training: Domain Fine-tuning

Take the translation of medical documents as an example:

Domain fine-tuning based on internal high-quality text can effectively improve the model performance of specific tasks.

Before fine-tuning: limitations of the general model

The understanding of professional terms and context is insufficient, and the accuracy and fluency of translation need to be improved.

After fine-tuning: the birth of domain experts

Accurately grasp industry terms and expression habits, and significantly improve translation quality and work efficiency.



AI Engineering - Model Training: Reinforcement Learning Alignment

Processing noisy data: Reward training based on TER



In the translation task, the **Translation Editing Rate (TER)** is used as a reward signal to guide the model to learn on noisy data and improve its robustness.

Handling specific negative cases: DPO(Direct Preference Optimization)



To solve the problem of inconsistent terms and untranslated words, construct negative examples, and use DPO method to directly optimize model preferences and align human expectations.



AI Engineering - MCP for Agent

MCP (model context protocol) defines three types of core semantic interfaces on top of standard rest, so that agents from different suppliers can discover, call and combine with each other.



Query CDISC Library

Agents understand and interpret complex CDISC standards.



Text Translation MCP

Wrapping Translation function to MCP



Data Search MCP

Wrap semantic search and key word search to MCP



AI Engineering - Data Query Platform based on ElasticSearch



Semantic retrieval: understand the query intention and return relevant results.



Keyword search: accurate matching, fast positioning.

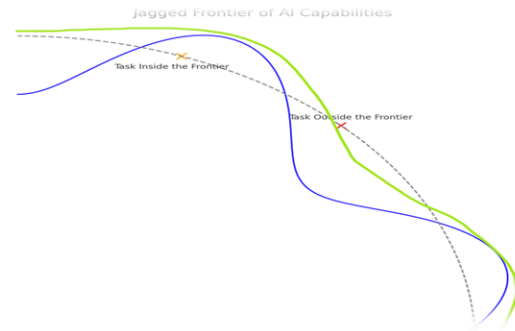


Hybrid search: combine multiple ways to provide the best query experience.



Summary

- ✓ **GenAI is reshaping biopharma;** clinical trials are the next transformation frontier, unlocking faster, smarter drug development.
- ✓ **Success hinges on three essentials:** high-quality, fit-for-purpose data; domain-specific customization; continuous, rigorous iteration and human verification;
- ✓ **Disciplined engineering and cross-functional collaboration can translate promising POCs into reliable and scalable production solutions -** turning data into clinical impact and accelerating time-to-market.



Moore's Law provides exponential growth in...

- Storage capacity per dollar
- Analytical bandwidth per dollar
- Data production per dollar

Moore's Law provides near-zero growth in...

- Labor pool
- Individual programming ability
- Individual visual acuity
- Hours in the day

Possible Solutions?

- Better automation
- Higher-level abstractions for humans
- Combinations of the two

Thank You



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
✉ office@phuse.global
🌐 phuse.global

Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)
📘 [/phusebook](https://facebook.com/phusebook)
▶ [/phusetube](https://youtube.com/phusetube)
🌐 [/company/phuse](https://linkedin.com/company/phuse)