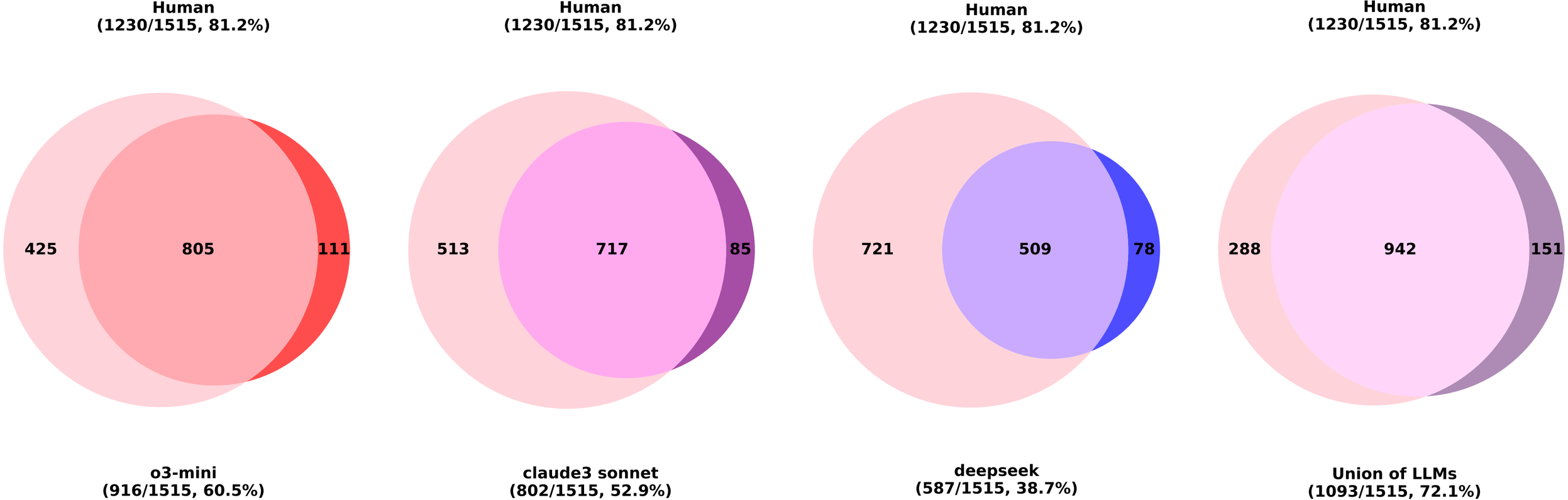
Human
(1230/1515, 81.2%)
425
805
111
o3-mini
(916/1515, 60.5%)
Human
(1230/1515, 81.2%)
513
717
85
claude3 sonnet
(802/1515, 52.9%)
Human
(1230/1515, 81.2%)
721
509
78
deepseek
(587/1515, 38.7%)
Human
(1230/1515, 81.2%)
288
942
151
Union of LLMs
(1093/1515, 72.1%)