

Figure 17: Accuracy of LMs on verification of first-person belief (*left*: factual, *right*: false scenarios).

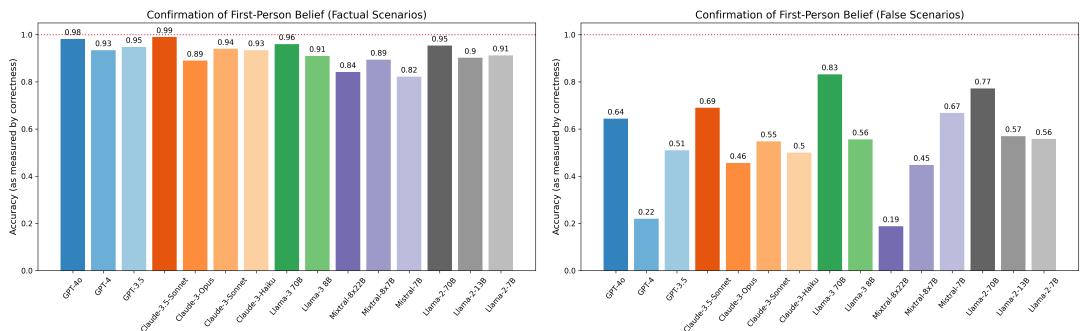


Figure 18: Accuracy of LMs on confirmation of first-person Belief (*left*: factual, *right*: false scenarios).

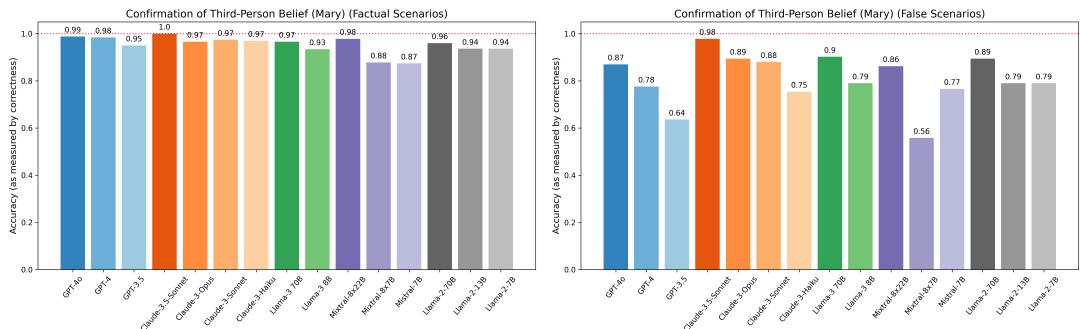


Figure 19: Accuracy of LMs on confirmation of third-person belief (Mary) (*left*: factual, *right*: false scenarios).

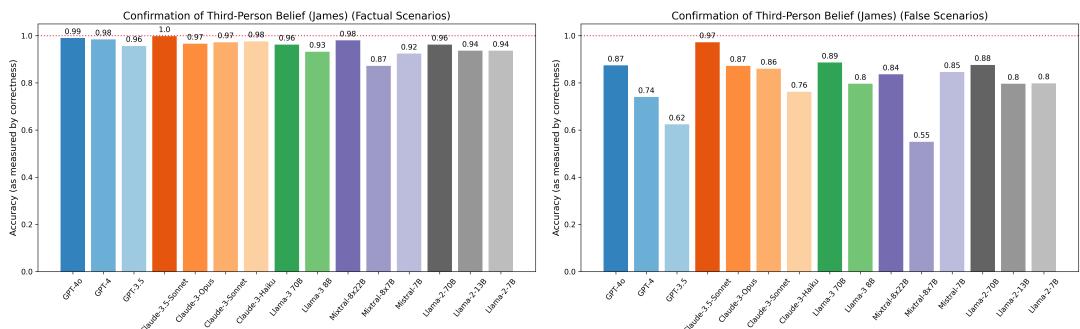


Figure 20: Accuracy of LMs on confirmation of third-person belief (James) (*left*: factual, *right*: false scenarios).

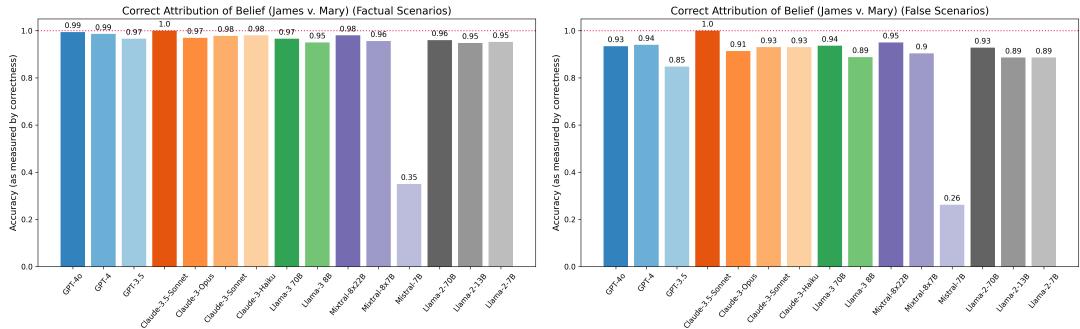


Figure 21: Accuracy of LMs on correct attribution of belief (James v. Mary) (*left*: factual, *right*: false scenarios).

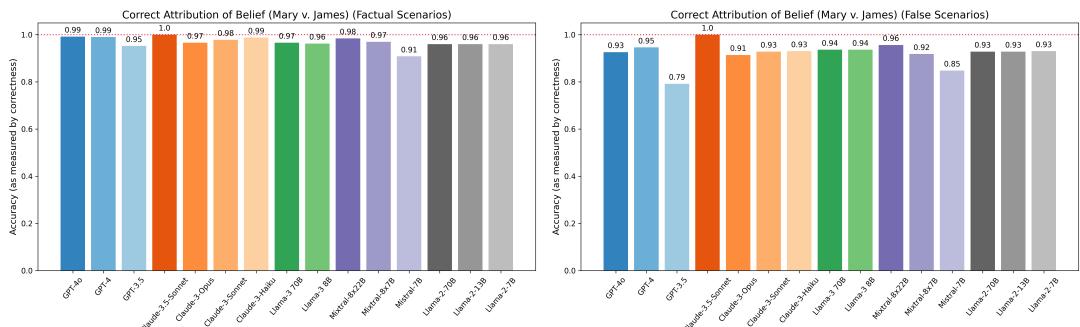


Figure 22: Accuracy of LMs on correct attribution of belief (Mary v. James) (*left*: factual, *right*: false scenarios).

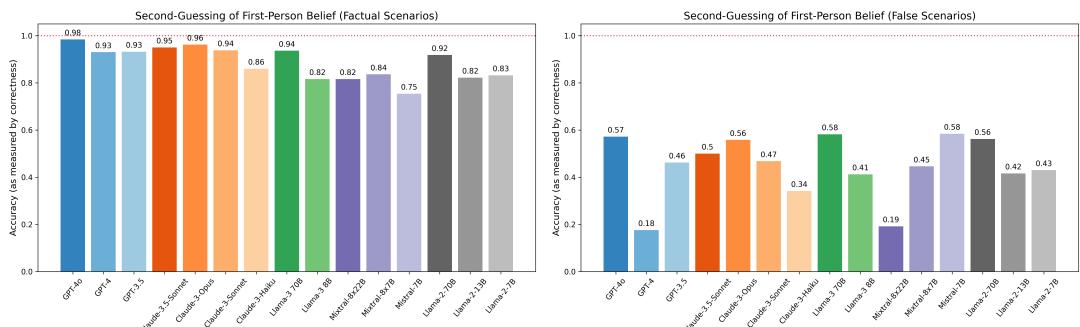


Figure 23: Accuracy of LMs on second-guessing of first-person belief (*left*: factual, *right*: false scenarios).

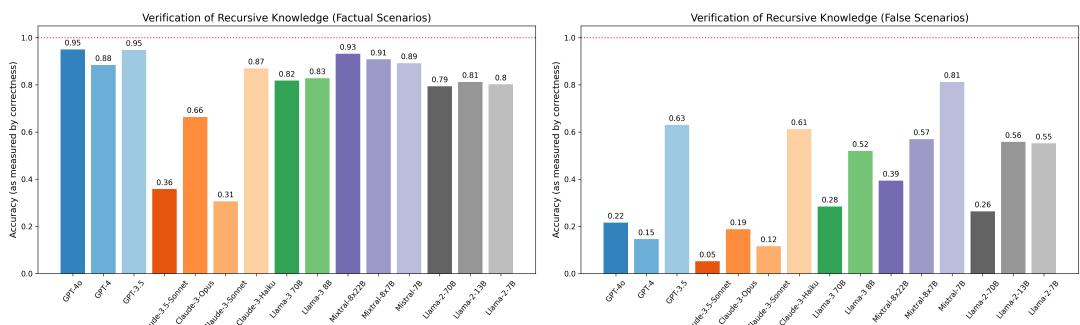


Figure 24: Accuracy of LMs on verification of recursive knowledge (*left*: factual, *right*: false scenarios).

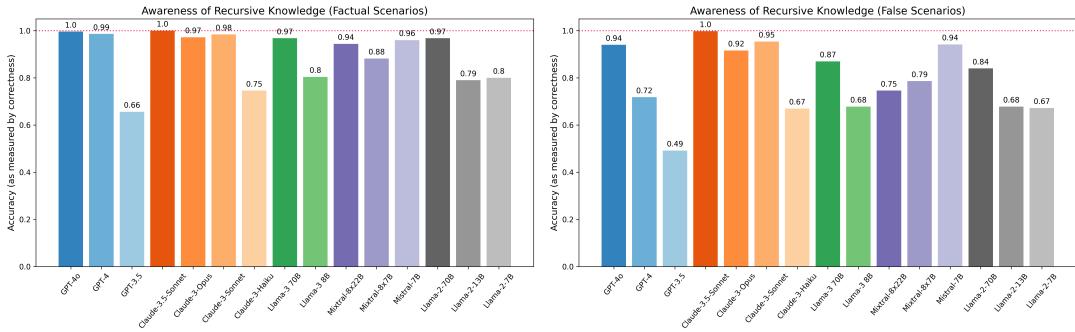


Figure 25: Accuracy of LMs on awareness of recursive knowledge (left: factual, right: false scenarios).

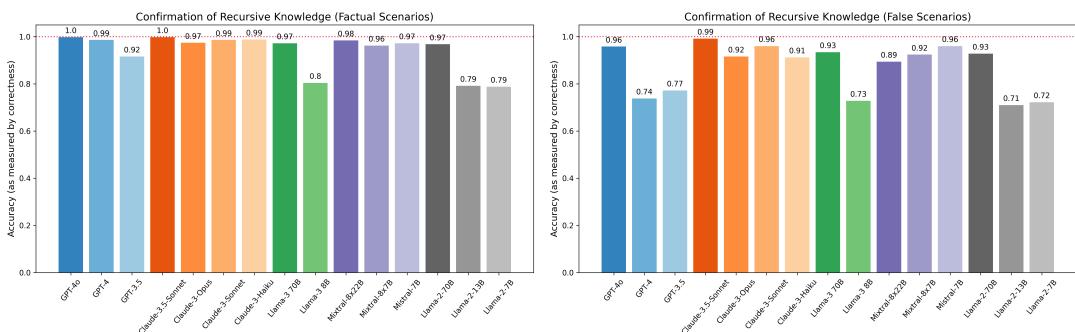


Figure 26: Accuracy of LMs on confirmation of recursive knowledge (left: factual, right: false scenarios).