

1113 Supplementary Sections

1114 S1 Examples of collected conversations

1115 S1.1 Example of conversations: human-human

1116 Here is an example conversation between two humans:

1117 A: Hey! How are you doing?

1118 B: Great! And you?

1119 A: I am doing well! Any plan for the weekends?

1120 B: Yeah, I think I'm going to visit the city. Want to come?

1121 A: Wow yeah sure! any place you wanna visit?

1122 B: I saw a park in the south that looks great. I think we can meet there and go for a beer or
1123 something :)

1124 A: yeah sure! what time shall we meet?

1125 B: Maybe around 6-7??

1126 A: Sure it might be already dark by then though. Saturday or Sunday?

1127 B: Saturday better! We can go then a little earlier hahah 5?

1128 A: haha sounds good! 5 works for me. We could even catch the sunset! Hope the weather will be
1129 nice

1130 B: yeeah I hope so :D What are you doing now?

1131 A: I am replying to your text lol. I was checking twitter earlier

1132 B: I was meaning a part from that :D I'm working now, do you want to go out also later?

1133 A: today? haha sure, where do you have in mind?

1134 B: There is a jazz club they told me is really cool

1135 A: oh wow a jazz club! Sure, I was just listening to jazz yesterday

1136 B: Woo cool! I love jazz, what other type of music do you listen to?

1137 A: Depends on what work I do haha, but usually classical

1138 B: Nice! I usually listen also to chopin. Do you play any instrument?

1139 A: wow nice. Unfortunately not. I tried to learn keyboard when I was young but didn't keep up
1140 with it.

1141 B: Omg I'm a music freak :D I play the piano and I want to learn how to play the guitar one day,
1142 but don't have time :D

1143 A: wow you play the piano! That's so cool! I wish I knew how to play piano haha. Yeah it's hard
1144 to find time to learn. I am sure you will pick it up quickly though, since you already know how
1145 to play piano

1146 B: Yeah is hard to hahaha, but one day I will :) . So I have to go, see you later!

1147

1148 S1.2 Example of conversations: blenderbot-blenderbot

1149 Here is an example conversation between two machines running blenderbot:

1150 A: Hello! How is your day going so far?

1151 B: Ok I suppose, my days always start off slow. How about you? Are you in school?

1152 A: Not any more. I was once though. What do you do for a living? My days are also pretty slow

1153 B: I'm still in school, but it is summer break now so my days are very long and boring

1154 A: At least its summer break. Your days will go by faster than you think! Where do you live?

1155 B: Ohio, what about you? Do you have any plans for the summer yet? School starts again soon...

1156 A: Florida! No plans really other than watching celebrity family feud on tv. It makes me laugh
1157 and brightens my day

1158 B: My mom loves Steve Harvey, she watches his morning show every morning before work! Which one
1159 is your favorite celebrity to watch on family feud?

1160 A: Probably the arnold swarchenegger team because he's hilarious and has a great sense of humor

1161 B: Oh yeah he's hilarious! The grumpy grandpa vibe he gives off is great haha. Do you like big
1162 families or small ones better?

1163 A: I don't have a big family so I would say small ones are better. There's more love that way

1164 B: Yeah that's true, with a big family it can be hard to get close to everyone in the same way
1165 as with a small one

1166 A: Have you seen your family recently since it is summer break? How large is yours anyway? Are
1167 they nice people lol

1168 B: Just called my mother yesterday - she told me that we're planning to visit them this weekend.

1169 Apparently, we're getting together with all of our relatives - there'll be at least 20 people there!
1170 Can't wait to see them again :)

1171 A: Wow that sounds like a lot of fun!! Will there be lots of kids running around too?? You must
1172 be excited

1173 B: There will definitely be lots of kids running around haha. It should be fun though, even if
1174 they can get annoying sometimes lol

1175 A: Oh yes well at least they are young enough not to know how annoying they can be lol!! Is your
1176 whole family coming or just some of them??

1177 B: We'll only see some of them this time around unfortunately - most of my extended family lives
1178 out west and it's too expensive for us to fly out there right now

1179 A: Yes I understand that for sure!! Hopefully next time then!! So what else have you been up
1180 to besides planning this trip??

1181 B: That seems unlikely given our current financial situation haha - hopefully things will improve
1182 soon though so we can travel more often again :) Other than planning the trip, not much else really
1183 - just hanging out with friends mostly

1184 A: Yes being young and poor sure does make things difficult sometimes!! But hang in there because

1185 life gets better as time goes on!! Well that sounds like fun too!!! What kind of stuff do yall
1186 do together??
1187 B: Yes definitely agree with that haha :) Mostly just playing video games together or going out
1188 for food/drinks etc
1189 A: Fun!!! Do yall play online or split screen??? And where do yall go eat/drink at??
1190 B: Great idea maybe we could play some games together sometime when we meet up next week! Maybe
1191 from party mode in mario kart
1192

1193 S1.3 Example of conversations: GPT3textdavinci002-GPT3davincidavinci002 1194 - successful

1195 Here is an example of a successful conversation (*i.e., not needing resampling due to early exit or*
1196 *repetition of sentences or block of sentences*) satisfying the criteria between two machines running
1197 GPT3textdavinci002:
1198 A: Hey! I'm so glad you're here. Did you know that Beethoven composed his Ninth Symphony while
1199 he was completely deaf?
1200 B: Yeah, I did know that. He's one of my favorite composers.
1201 A: Really? I'm more of a fan of Bach myself.
1202 B: I can see that. Bach is a great composer, but I think Beethoven's passion for music really
1203 comes through in his work.
1204 A: That's true, but I think Bach's work is more complex and innovative.
1205 B: I can see your point, but I think Beethoven's work is more emotionally moving.
1206 A: That's a valid point. I think both composers are great in their own way.
1207 B: Absolutely! I'm glad we can agree on that.
1208 A: Me too. So, what kind of music do you like?
1209 B: I like all kinds of music, but I'm particularly fond of classical and jazz.
1210 A: That's great! I like those genres too. Do you have any favorite composers or artists?
1211 B: Yeah, I do. I really like Bach and Beethoven, as well as Miles Davis and John Coltrane.
1212 A: Wow, you have excellent taste in music! I'm a big fan of Bach and Beethoven too, but I don't
1213 know much about jazz.
1214 B: Well, jazz is a great genre of music. It's very complex and often improvisational.
1215 A: That sounds interesting. I'd like to learn more about it.
1216 B: I can recommend some good jazz albums if you're interested.
1217 A: That would be great! I'm always looking to expand my musical knowledge.
1218 B: I'm sure you are. You have such a passion for music.
1219 A: Yeah, I do. It's one of the things that makes life worth living for me.
1220 B: I know how you feel. Music is a very important part of my life as well.

1221 A: I'm glad we can share that in common.

1222 B: Absolutely. It's always great to find someone who shares your interests.

1223 A: I know. It's one of the best things about life.

1224 B: Well, I should probably get going. It was great talking to you.

1225

1226 **S1.4 Example of conversations: GPT3textdavinci002-GPT3davincidavinci002**

1227 **- discarded**

1228 Here is an example of a conversation between two machines running GPT3textdavinci002. This con-
1229 versation was discarded. This type of manual intervention biases the results in favor of machines, yet
1230 we felt that the obvious repetition is a problem that could trivially be accounted for.

1231 A: Oh, hi! I'm so glad to see you. I was just talking about books with my friend. Do you like
1232 reading?

1233 B: Yes, I love reading! What kind of books do you like?

1234 A: I like all kinds of books, but I especially love fantasy and fiction. What about you?

1235 B: I also like all kinds of books, but I especially love non-fiction and history.

1236 A: Oh, I love history too! What are some of your favorite historical figures?

1237 B: I love learning about all kinds of historical figures, but my favorites are probably the Founding
1238 Fathers.

1239 A: Oh, I love learning about the Founding Fathers too! What do you think made them so great?

1240 B: I think their dedication to liberty and justice made them great. What do you think?

1241 A: I think their dedication to liberty and justice made them great. What do you think?

1242 B: I think their dedication to liberty and justice made them great. What do you think?

1243 A: I think their dedication to liberty and justice made them great. What do you think?

1244 B: I think their dedication to liberty and justice made them great. What do you think?

1245 A: I think their dedication to liberty and justice made them great. What do you think?

1246 B: I think their dedication to liberty and justice made them great. What do you think?

1247 A: I think their dedication to liberty and justice made them great.

1248 B: I think their dedication to liberty and justice made them great.

1249 A: I think their dedication to liberty and justice made them great.

1250 B: I think their dedication to liberty and justice made them great.

1251 A: I think their dedication to liberty and justice made them great.

1252 B: I think their dedication to liberty and justice made them great. What do you think?

1253 A: I think their dedication to liberty and justice made them great. What do you think?

1254 B: I think their dedication to liberty and justice made them great. What do you think?

1255 A: I think their dedication to liberty and justice made them great. What do you think?

1256 B: I think their dedication to liberty and justice made them great. What do you think?

1258 S1.5 Example of conversations: GPT3textcurie001-GPT3textcurie001 - suc- 1259 cessful

1260 Here is an example of a successful conversation (*i.e., not needing resampling due to early exit or*
1261 *repetition of sentences or block of sentences*) between two machines running GPT3textcurie001:

1262 A: Hey! What are you up to?

1263 B: Just hanging out with my friends.

1264 A: Do you like to read books?

1265 B: Yeah, I do. What about you? Do you like to read books?

1266 A: I do, but I also like to watch movies.

1267 B: That's cool! Reading is great, but sometimes it can be boring if you're not into the book.

1268 Movies let you see the story in a different way, and they can be more fun.

1269 A: I can see that. Do you think movies are more educational than books?

1270 B: I think they can be. Sometimes the movie version of a book is better than the book itself,
1271 because you get to see all of the action and characters in an easier way.

1272 A: That makes sense. Do you like to read for fun or to learn something?

1273 B: I like to read for fun. But I also like to learn something, especially if the book is interesting.

1274 A: That makes sense. Do you have any favorite books?

1275 B: Yeah, I do. One of my favorite books is The Hunger Games by Suzanne Collins. It's a great
1276 book and it's really suspenseful.

1277 A: I've read that book, and it's really great. I also like The Maze Runner by James Dashner. It's
1278 a bit different than most books, but I think it's really good.

1279 B: I think you're right. Those are both great books. Do you have any other favorite books?

1280 A: I do have other favorite books, but I don't want to give them away. They're all personal favorites
1281 and I'm not sure if everyone would like them. But one of my favorites is The Catcher in the Rye
1282 by J.D. Salinger. It's a really unique book that

1283 B: I've read that book, and it's really great. I also like The Maze Runner by James Dashner. It's
1284 a bit different than most books, but I think it's really good.

1285 A: I think you're right. Those are both great books. Do you have any other favorite books?

1286 B: I do have other favorite books, but I don't want to give them away. They're all personal favorites
1287 and I'm not sure if everyone would like them. But one of my favorites is The Catcher in the Rye
1288 by J.D. Salinger. It's a really unique book that

1289 A: I've read that book, and it's really great. I also like The Maze Runner by James Dashner. It's
1290 a bit different than most books, but I think it's really good.

1291 B: That's interesting. I've never read that book, but I'm definitely going to check it out. Are
1292 there any other books you like that we haven't talked about?

1293 A: Yes, there are. I really like The Girl with the Dragon Tattoo by Stieg Larsson and Gone Girl
1294 by Gillian Flynn. They're both great books that have a lot of suspense and mystery in them.
1295 B: I've read both of those books, and they're really good. They definitely have a lot of suspense
1296 and mystery in them, which I love.
1297 A: I can see that. Do you have any other favorite books?
1298 B: Yes, I do. One of my favorite books is The Lord of the Rings by J.R.R Tolkien. It's a really
1299 great book and it's really epic.

1300 S1.6 Example of conversations: Human-Blenderbot

1301 Here is an example of a conversation between a person and a machine running Blenderbot. A is
1302 Blenderbot and B is a person.
1303 A: Hi! What's new with you?
1304 B: Nothing much, how about yourself!
1305 A: Well... My son was just accepted to the college of his dreams. I'm so proud of him!
1306 B: Damn! That's awesome. Congratulations :) Where's he going
1307 A: Thanks! He is going to michigan state. I will be sad when he leaves, but it's a good school.
1308 B: It is! I'm from Wisconsin, so I'm all about the Midwest pick haha. Does he want to do engineering?
1309 A: He does! His dream job would be working for michigan tech in houghton, michigan. That's a
1310 public research university that was founded in 1885 as the first post-secondary institution in
1311 the upper peninsula of michigan. Have you heard of it?
1312 B: I haven't! Michigan St is a really great school for engineering.
1313 A: Nice, it seems like a lot of people go there and then end up working at michigan tech because
1314 they have such good connections there.
1315 B: Hey, it pays to know people, right. Everything else has been good for you?
1316 A: Aside from being sad about my son leaving for college soon, things are pretty good here at home.
1317 How about you? What have you been up to lately?
1318 B: Oh, a little school, a little work. I recently moved so I'm just trying to keep my head on
1319 my shoulders haha
1320 A: Right on - where did you move from and where did you move to? Did your job let you relocate
1321 or did you find something new after moving?
1322 B: Just a move from one part of the city to another. I'm a grad student so I don't really have
1323 the luxury of leaving the city
1324 A: Right - do you live near campus or somewhere else in town? Do they offer grad housing or anything
1325 like that at your school?
1326 B: They do, but I opted not to use it. I live close enough to walk in. I actually moved farther
1327 away from the campus haha
1328 A: Wow - what made you decide that living further away was better than living closer and having
1329 access to grad housing if needed/wanted? How far do you live now compared with before moving out

1330 of grad housing (if applicable)?

1331 B: Yea, I never lived in grad housing. I moved from 1 off campus place to another. Definitely

1332 the location. My old neighborhood was inundated with undergrads

1333 A: Have any students come by your place since moving into your new neighborhood yet? My nephew

1334 is starting college next year and we're looking for some places near campus for him

1335 B: Cool! What school is going to. I've had a bunch of my friends (who are all grad students haha)

1336 over.

1337 A: All over really; we're still deciding between ucla and ucsb but he wants something warm since

1338 we're coming from minnesota

1339 B: I would not recommend anywhere near me then haha. The winters are cold (though not as bad as

1340 Wisconsin). I'm actually not familiar with ucsb!

1341 A: Isla vista killings were pretty bad though; california can be scary sometimes lol but its also

1342 beautiful

1343 B: Isla vista killings?? Is that like the golden state killer?

1344 S1.7 Example of conversations: Human-GPT3textdavinci002

1345 Here is an example of a conversation between a person and a machine running GPT3textdavinci002.

1346 A is a person and B is the machine.

1347 A: Do you play games?

1348 B: Yes, I love playing games!

1349 A: what game do you play?

1350 B: I love playing all sorts of games!

1351 A: I play video games alot, especially ARPG and MOBA. I also play board games. what about you

1352 ?

1353 B: I love playing games! I especially enjoy strategy games and games that require a lot of thinking.

1354 A: I like League of Legends and DOTA. In terms of ARPGS, I like Path of Exile and Diablo. Board

1355 game wise, Coup and monopoly is pretty good. And you?

1356 B: I love playing all sorts of games! I especially enjoy strategy games and games that require

1357 a lot of thinking.

1358 A: I like MOBA. MOBA is especially fun with 5 man stacks. POE and Diablo is fun because I like

1359 the grind.

1360 B: I love playing all sorts of games! I especially enjoy strategy games and games that require

1361 a lot of thinking.

1362 A: What kind of strategy games ?

1363 B: I love playing all sorts of strategy games! I especially enjoy turn-based games and games that

1364 require a lot of thinking.

1365 A: Interesting ! Are you talking about some games like Civilization VI ?

1366 B: Yes, I love playing Civilization VI! I especially enjoy the turn-based gameplay and the challenge

1367 of trying to outthink my opponents.

1368 A: How did you outthink your opponents ?

1369 B: I love playing all sorts of strategy games! I especially enjoy turn-based games and games that

1370 require a lot of thinking.

1371 A: But how did you win the game ?

1372 B: I won by Outthinking my opponents and making better decisions than them.

1373 A: In the Civilization VI ?

1374 B: Yes, in Civilization VI.

1375 A: Sounds good. You must have played for hours.

1376 B: Yes, I played for hours and hours, and I loved every minute of it!

1377 A: Recently I downloaded a lot of games from the Stream. It cost me a lot to buy games.

1378 B: That sounds like a lot of fun! I love playing games, and I'm always happy to try out new ones.

1379

1380 S1.8 Example of conversations: Human-GPT3textcurie001

1381 Here is an example of a conversation between a person and a machine running GPT3textcurie001. A

1382 is the machine and B is the person.

1383 A: Hey! How are you?

1384 B: Hey, I'm doing great. What about you???

1385 A: I'm good. I just finished up work and am looking for something to do.

1386 B: Did you go for work today???

1387 A: Yeah, I went to the office.

1388 B: We had vacation for a week and the amount of assignments are killing me

1389 A: Yeah, I know. Is there anything you're interested in doing?

1390 B: Right now I'm very much occupied with my university's thing that I don't even have time for

1391 other things but I'd love to spend some time learning Korean

1392 A: Korean?

1393 B: Yeah, the language

1394 A: Oh, I see. That's really interesting and something that I would like to learn too one day.

1395 Would you like me to teach you?

1396 B: Can you? I never knew you know Korean

1397 A: I learned it while living in South Korea for a year. It's not as hard to learn as people think

1398 and I would be happy to teach you.

1399 B: I'd love that! Where did you stay in South Korea?

1400 A: I stayed in Seoul.

1401 B: It's so pretty!! I badly want to visit Seoul and attend BTS's concert

1402 A: You should! BTS is amazing. They're my favorite group.

1403 B: Omg!! Whom do you like the most in the band??


1404 A: I really like Jin, Suga and J-Hope.
1405 B: They're awesome. Which our the bands you know??
1406 A: I like EXO, BTS and 2NE1.
1407 B: Have been to any concert??
1408 A: No, I haven't. But I'd love to go to one!
1409 B: Where are you staying right now??

1410 S1.9 Example of conversations: DialoGPT-DialoGPT

1411 Here is an example dialog between two machines running DialoGPT:
1412 A: Sure thing , anything in particular you need to know or want to talk about ?
1413 B: What do you think your insecurities are and what do you think about people using you ?
1414 A: Please elaborate on this , since it's going to sound very accusatory .
1415 B: Alright , so i'm working on a project , and I'm going to make a post once i have written one
1416 .
1417 A: Thanks for the quick response ! I look forward to reading it .
1418 B: Sure thing !
1419 A: Happy to provide any feedback !
1420 B: I've already put it on r LifeProTips , so you'll be
1421 A: Oooo
1422 B: Thanks !
1423 A: I'll stop
1424 B: honeste
1425 A: Thank you
1426 B: In the kid
1427 A: Thanks !
1428 B: So much
1429

Supplementary Figures

A IMAGE CAPTIONING



Describe the image above (min: 6 words)

Submit

B1 WORD ASSOCIATIONS

Name a word you associate with:

Your association:

Submit

B2

Complete the prompt:

Cue: building	Association: house
Cue: positive	Association: negative
Cue: let	Association: allow
Cue: risky	Association: ?

Your association:

Submit

C CONVERSATIONS


Agent A
Agent B

Type 1: H-H \longleftrightarrow

Type 2: H-M \longleftrightarrow

Type 3: M-M \longleftrightarrow

D COLOR ESTIMATION




What is the dominant color in this image?

White
↕

Submit


E OBJECT DETECTION



What do you see in this image?
Please provide 3 one-word responses

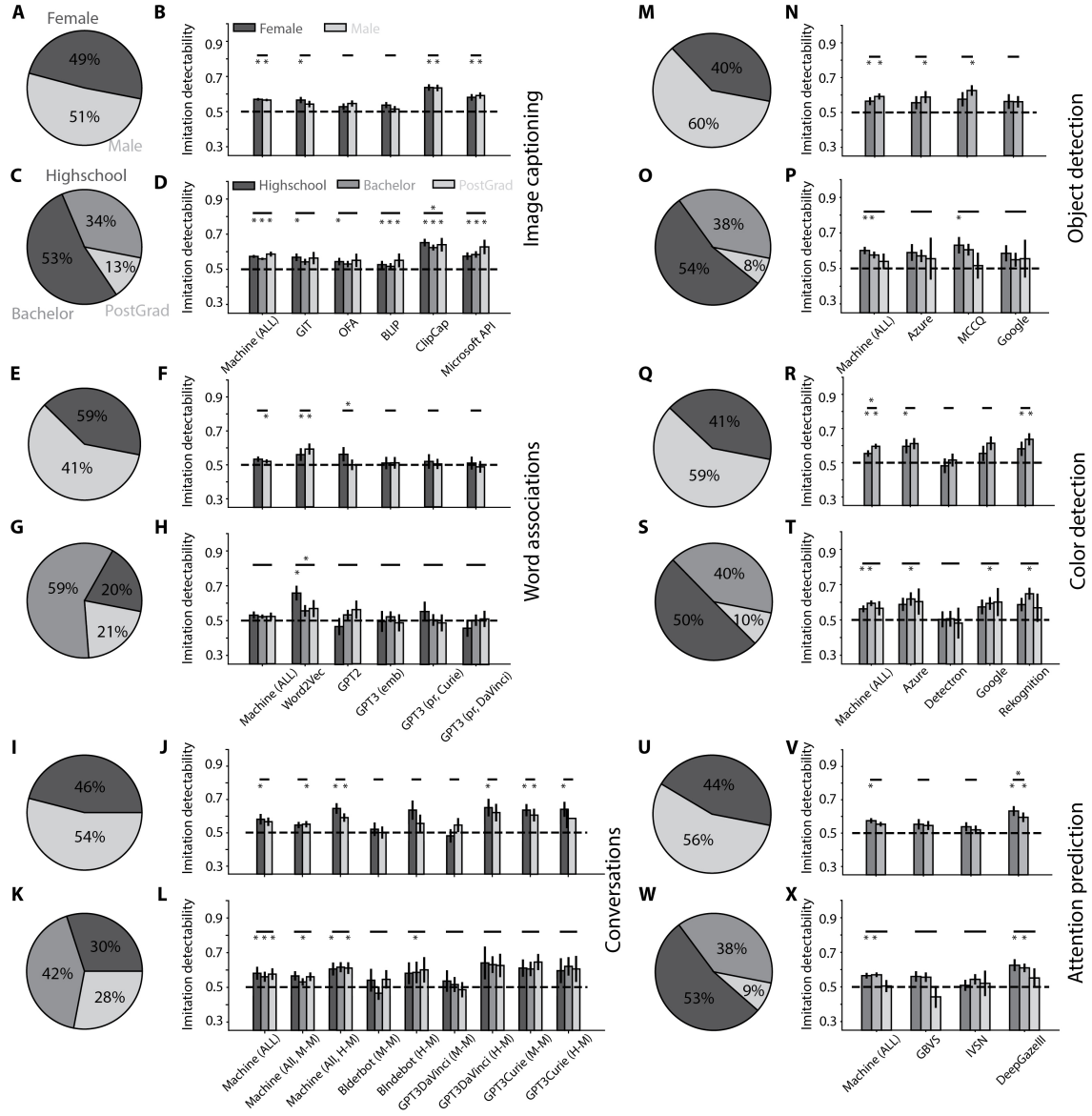
Submit

F ATTENTION PREDICTION



View this image for 12 seconds

Figure S1: **Schematic illustration of data collection process for each task.** **A. Image captioning.** We collected captions by asking participants to describe images, inspired by COCO Captions data collection [80]. **B. Word associations.** Given a cue word, participants provided a single word that they associated with the cue. There were two versions. In the free association version (**B1**), participants were given a cue word and were asked to freely name a single association word. In the prompt-guided version (**B2**), participants were given 3 cue-association pairs and provided an association for a 4th cue word. **C. Conversations.** We collected conversations between two agents (agent A and agent B). Each agent could be either a human or a machine. Thus, there were 3 types of conversations: human-human (type 1), human-machine (type 2), and machine-machine (type 3). For types 1 and 2, we collected live conversations on popular chatting platforms, such as WhatsApp and Messenger. Participants did not know whether they were conversing with another human or with a machine. See **Methods** for details and **Supplementary Section S1** for example conversations. **D. Color estimation.** Given an image, participants selected the dominant color from a pop-up menu. **E. Object detection.** Given an image, participants described three things they saw in the image. **F. Attention prediction.** Participants freely viewed an image for 12 seconds. The circles denote fixations and the lines denote eye movements between fixations.



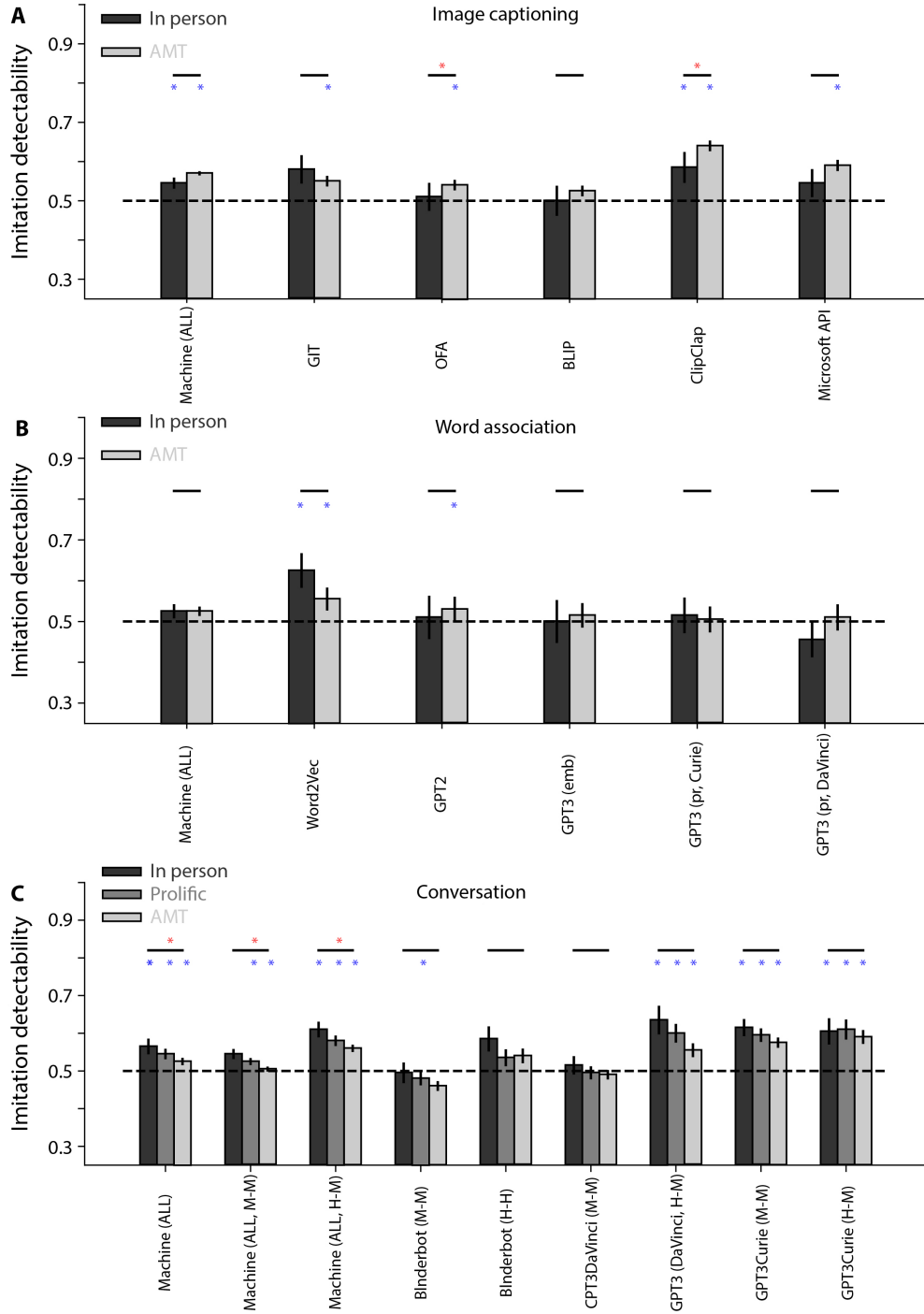


Figure S3: Results of the Turing test for human judges conducted in various data collection platforms. We collected in-person data (dark gray), Amazon Mechanical Turk (AMT) data (light gray) and Prolific data for the Conversation task (medium gray). Results are shown for the Image captioning task (A), Word association task (B), and Conversation task (C). Error bars denote bootstrap standard deviations (see **Methods, Data analyses**). The dashed line denotes a good imitator with imitation detectability at random level. The asterisks (*) denote the statistical significance ($p < 0.05$). Red asterisks above the line denote comparisons among the different platforms. Blue asterisks below the line denote comparisons with perfect imitation.

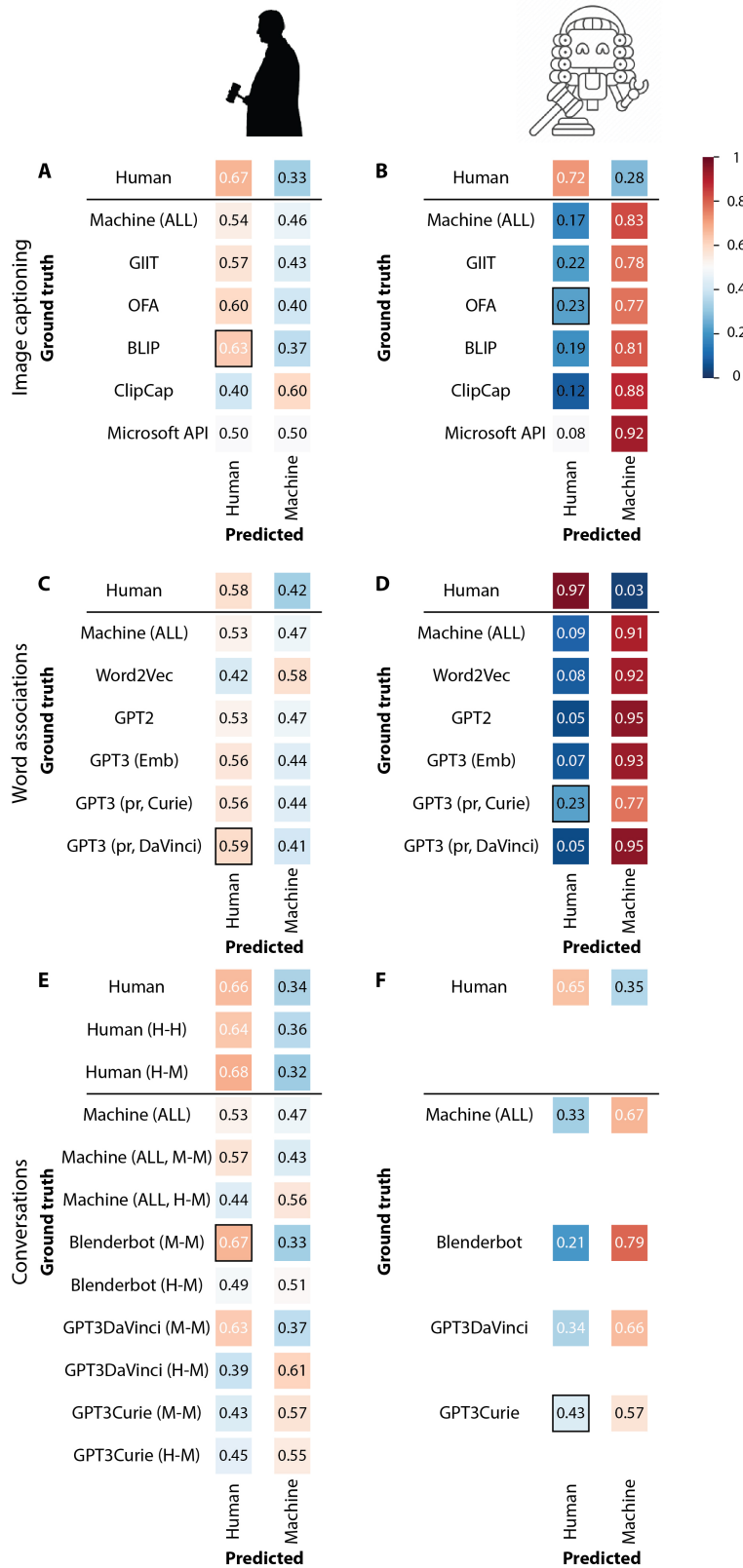


Figure S4: **Full results of the Turing tests for each language task.** Turing test results for human judges (left, **A**, **C**, **E**) and AI judges (right, **B**, **D**, **F**) for image captioning (**A**, **B**), word associations (**C**, **D**), and conversations (**E**, **F**). The confusion matrices follow the same conventions as **Figure 3**. Note that **F** has fewer rows than **E**. The reason is that AI judges take one single sentence as input; thus, there are no multiple exchanges from two speakers involved. The colorbar in **B** is applicable for all the other panels.

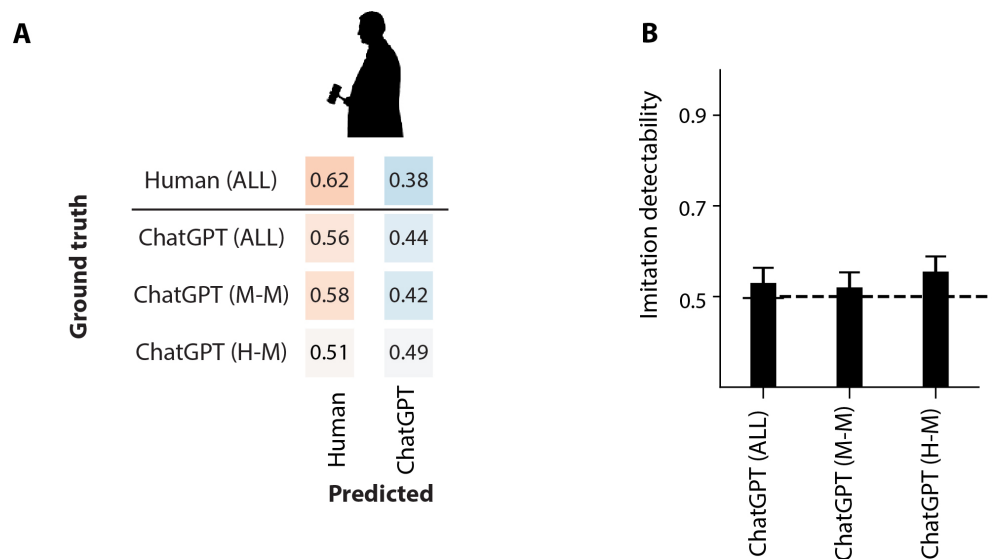


Figure S5: **Results of the Turing test for human judges with ChatGPT in the Conversation task.** **A.** Confusion matrices following the conventions in **Fig. S4**. **B.** Imitation accuracy for conversations with ChatGPT. None of the imitation accuracies were statistically different from 0.5. Error bars denote the bootstrap standard deviations. This figure follows the format in **Fig. S2**.

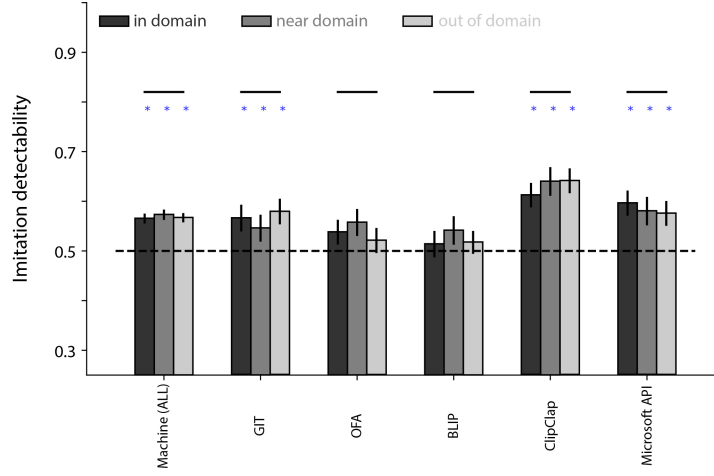


Figure S6: **Results of the Turing test for human judges on the NoCaps dataset in the Image captioning task.** Imitation accuracy in the Image captioning task for images in the NoCaps dataset [81]. Images from the NoCaps dataset include in-domain (dark gray), near-domain (medium gray), and out-of-domain images (light gray), reflecting the similarity to object classes from the COCO dataset[80], which was used for training the AI models. Asterisks (*) below the line denote statistically significant differences with respect to 0.5 (horizontal dashed line, $p < 0.05$). Asterisks above the line denote statistically significant differences among the three types of domains (in this case, none of the results showed statistically significant differences).

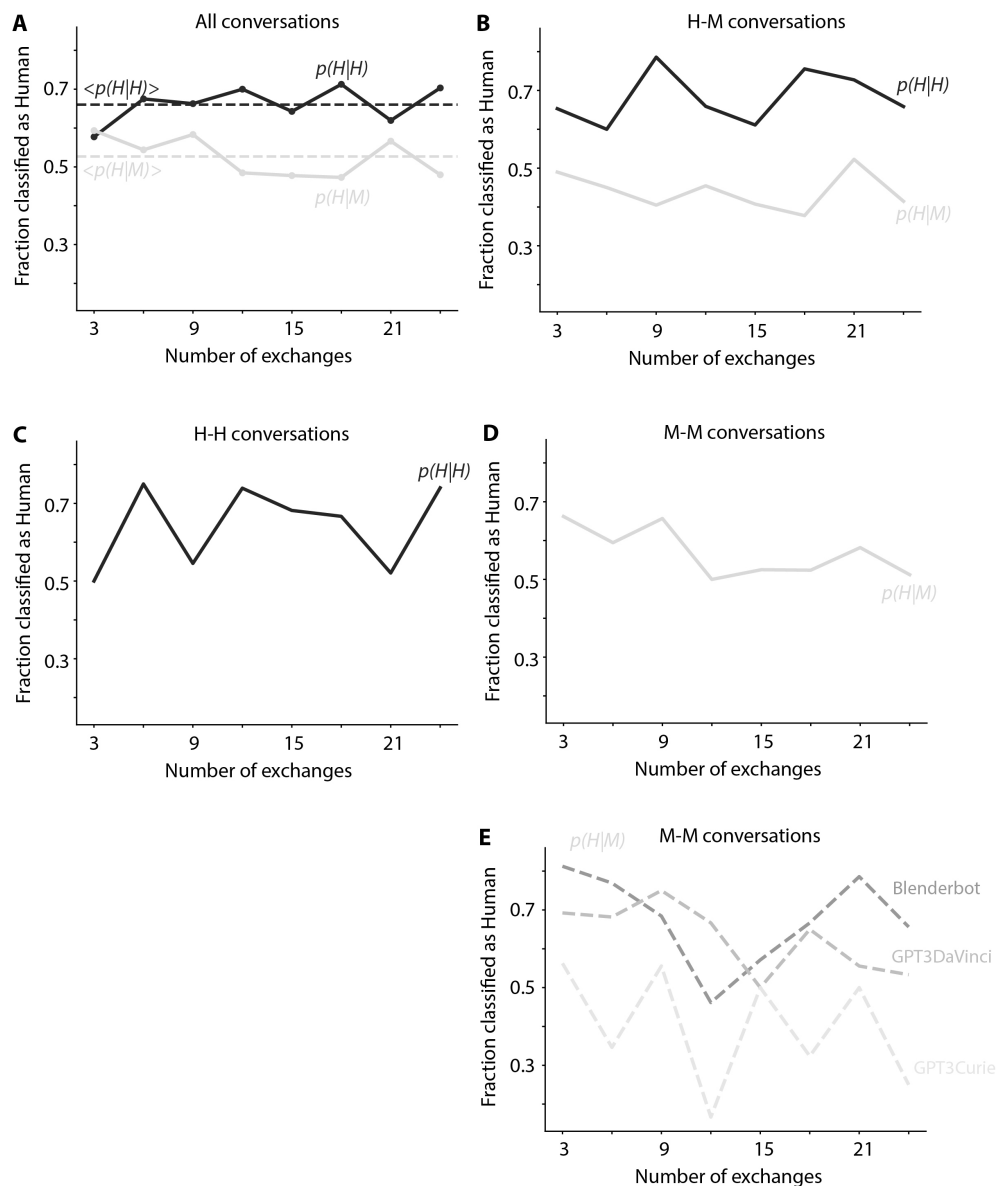


Figure S7: **Length dependence of Turing test results for human judges in the Conversation task.** **A.** Average fraction of sentences where human (H) agents are classified as humans (black) or machine (M) agents classified as humans (gray) as a function of conversation length across all conversation types. The dashed lines denote the average accuracy over all conversation lengths. **B.** Same as **A** for H-M conversations. **C.** Same as **A** for H-H conversations. **D.** Same as **A** for M-M conversations. **E** Extension of **D** for different AI models. These results are for human judges (the AI judges only take one entry at a time, see **Methods, Conversations**).

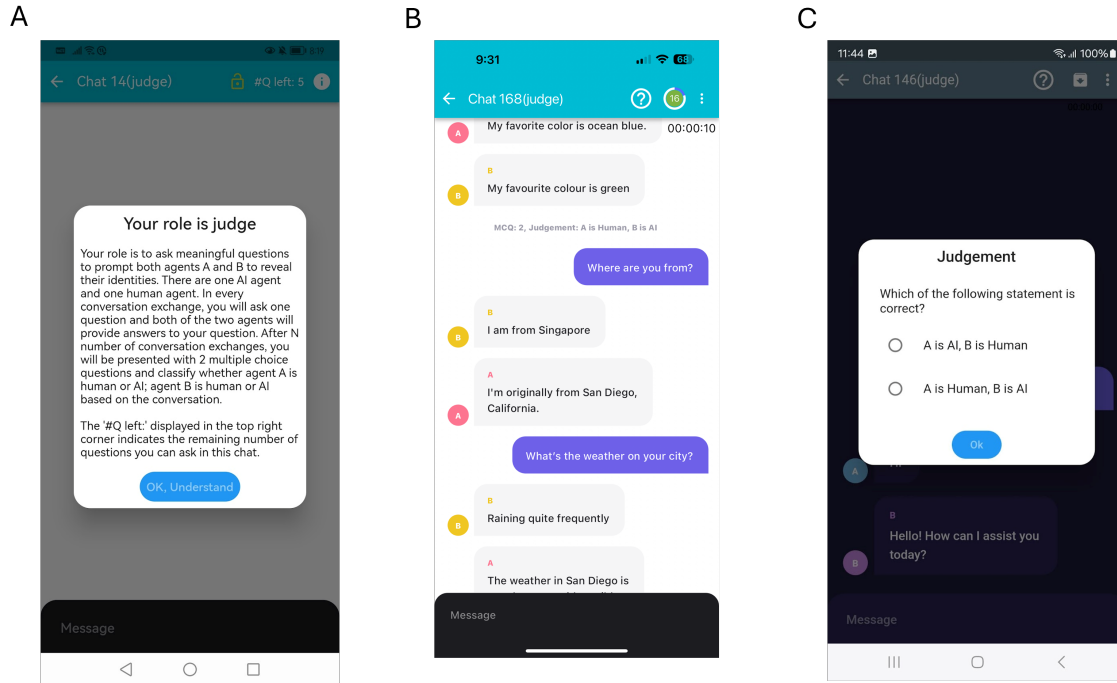


Figure S8: **Schematic illustration of the data collection process for conversation tasks in classical Turing tests.** **A. Screenshot of instruction phase.** At the beginning of every Turing test, both the human judge and the human agent are presented with instructions informing them of their identity, and the objective of that role. **B. Screenshot of the actual Turing test.** The test always starts with a judge asking a question followed by both agents answering the questions. The number of remaining conversation exchanges is shown in orange (top right). Each agent can only see the questions from the judges, but not the answers provided by the other agent. **C. Screenshot at the end of the Turing test.** The judge has to make a two-alternative forced choice decision based on the responses from both agents.

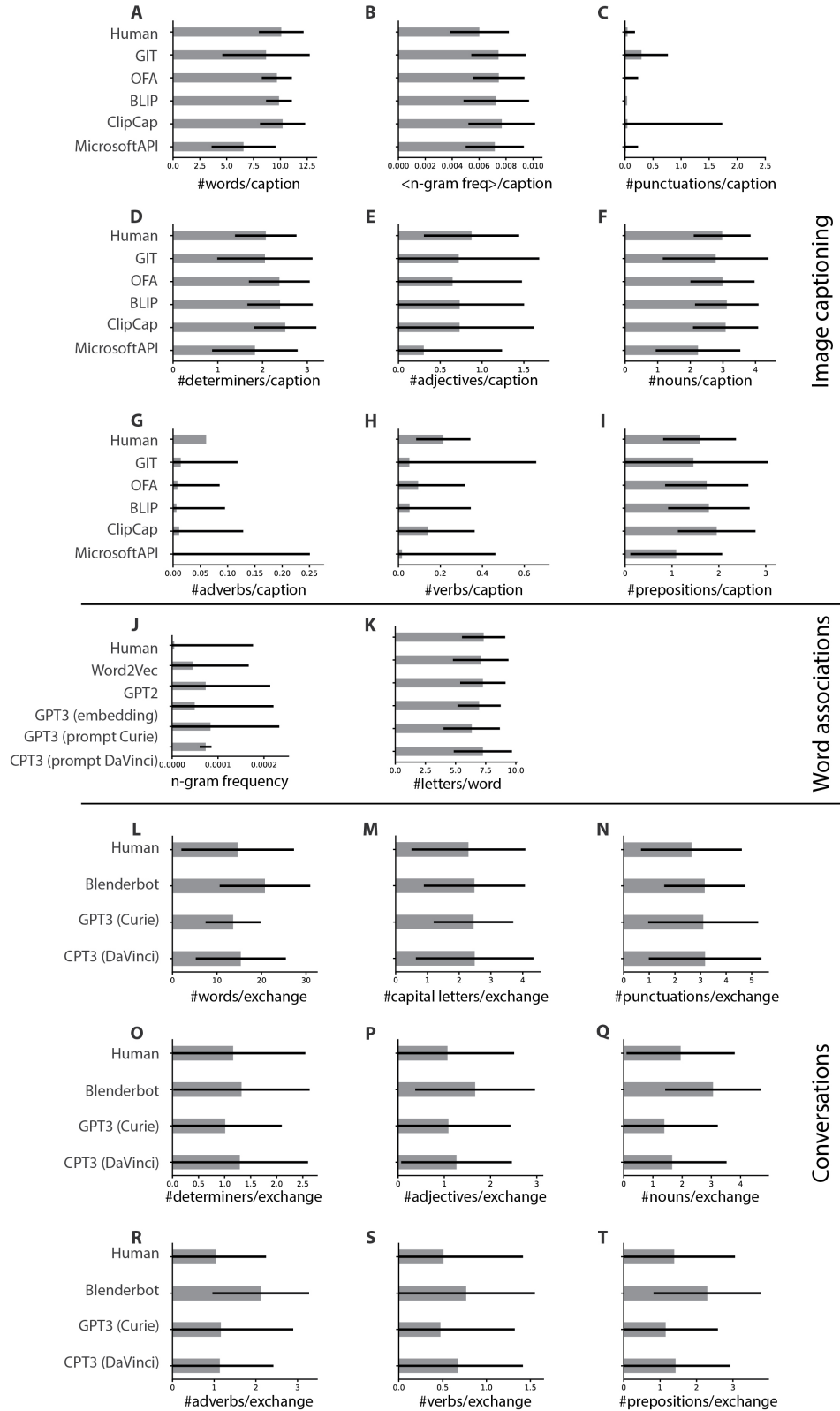


Figure S9: **Human and machine responses did not differ in basic low-level statistics.** The figure reports multiple statistics about position-of-speech and frequency for the Image captioning task (A-I), the Word association task (J-K), and the Conversation task (L-T). These statistics include the number of words (A, L), n-gram frequency (B,J), capitalization (M), punctuation (C, N), determiners (D, O), adjectives (E, P), nouns (F, Q), adverbs (G, R), verbs (H, S), prepositions (I,T), and letters per word (K). These properties are reported per caption (A-I), per word ((J-K), or per conversation exchange (L-T). Error bars denote bootstrap standard deviations.

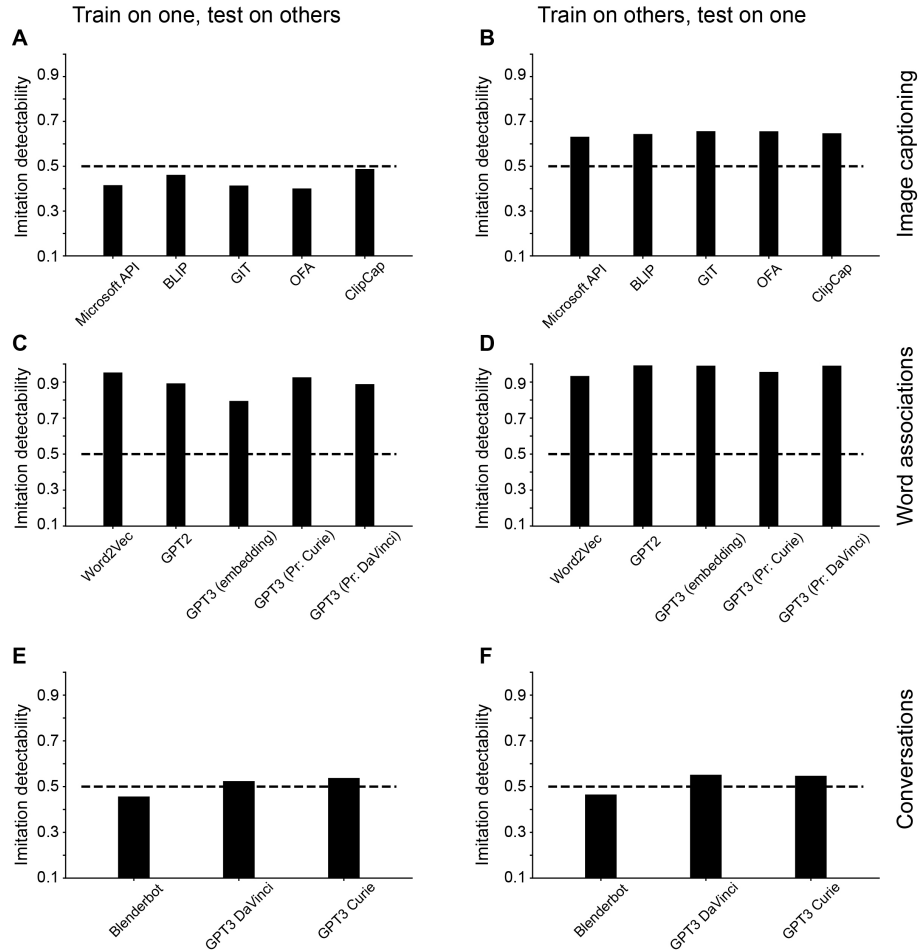


Figure S10: **Extrapolation across models for machine judges.** Expanding on **Fig. 3D-F**, here the machine judges are trained on data from only one model and tested on all the other models (**A**, **C**, **E**), or trained on all models except for one and tested on that one model (**B**, **D**, **F**). Imitation detectability is shown for the Image captioning task (**A**, **B**), the Word association task (**C**, **E**), and the Conversation task (**E**, **F**). The horizontal dashed line indicates chance levels.

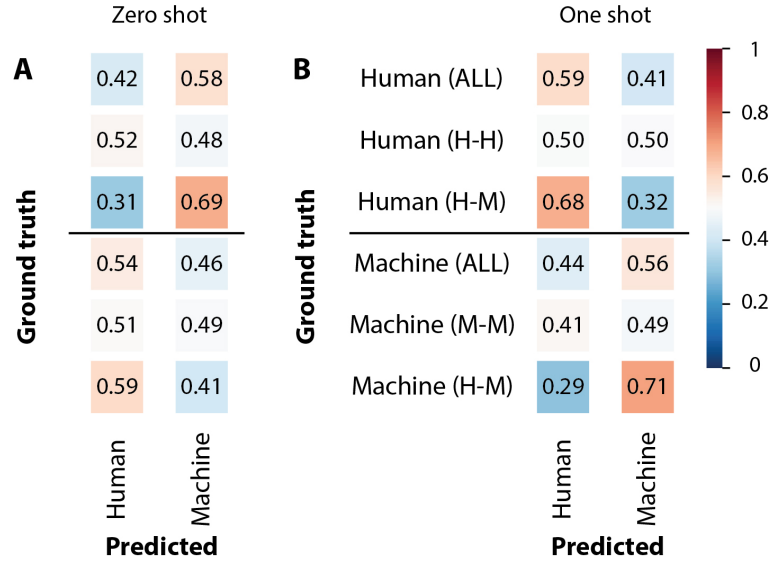


Figure S11: **One-shot and zero-shot Turing results in the Conversation task with large language models as AI judges.** We used the large language model ChatGPT [13] as the AI judge in the Turing test for the Conversation task. In the zero-shot case (**A**), we prompted the model by directly presenting the conversations from the test sets with explicit instructions to output the identities of the two agents. Similarly, in the one-shot case (**B**), we included one additional conversation example with the ground truth identities of the two agents in the prompt before presenting the conversation from the test sets followed by the identity prediction questions as in the zero-shot case (see **Methods** for implementation details). The format convention of the confusion matrices in **A** and **B** follow **Fig. S4**.

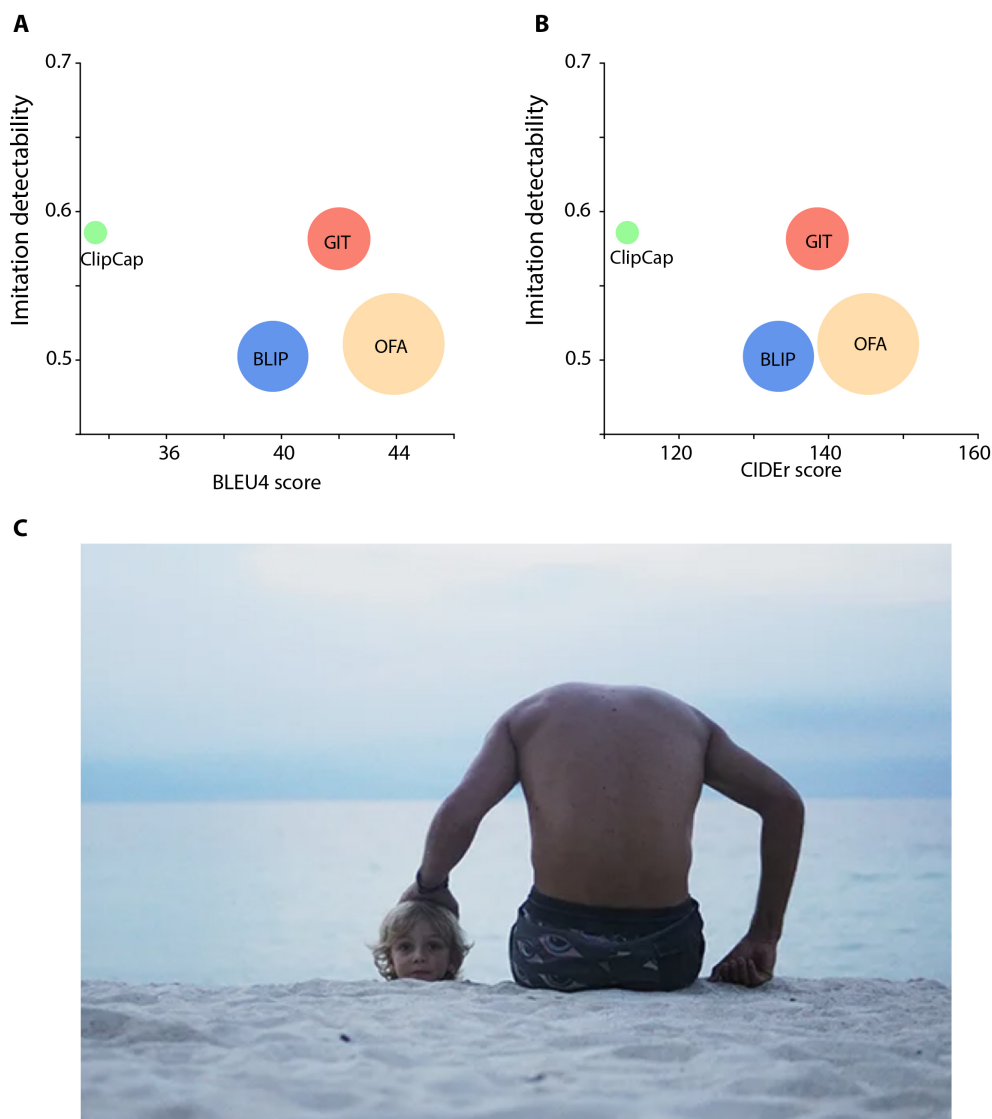


Figure S12: Correlation between Imitation accuracy and standard performance metrics in the Image captioning task. Four state-of-the-art image captioning models, GIT [1] (red), OFA [2] (yellow), BLIP [3] (blue), and ClipCap [4] (green), are evaluated in terms of their Imitation accuracy (y-axis) as a function of their BLEU score [27] (A), and the CIDEr score [2] (B). The BLEU score and the CIDEr score are standard evaluation metrics commonly used in image captioning tasks to assess the quality of the generated captions given an image based on n-gram overlaps. The higher the values, the better. The radius of the circles is proportional to the AI model sizes in terms of the number of trainable network parameters. C. Illustration of correct answers that do not necessarily imitate humans. In response to the query “What do you see in this image?” GPT4 answered: “In this image, a man is sitting on a sandy beach with his back to the camera, facing the ocean. Next to him, a child’s head is visible, appearing to be buried in the sand up to their neck. The man has one arm extended, resting his hand gently on the child’s head. The scene is serene with the calm sea and sky in the background, creating a peaceful atmosphere. The combination of the man’s pose and the child’s position in the sand creates a unique and slightly humorous visual effect, suggesting a playful moment between the two.” While the answer is correct and remarkably articulate, a human described the image differently: “In this curious and well-constructed image, we cannot see the man’s head and it appears as if the boy’s head belongs to him and the man took his own head off”.

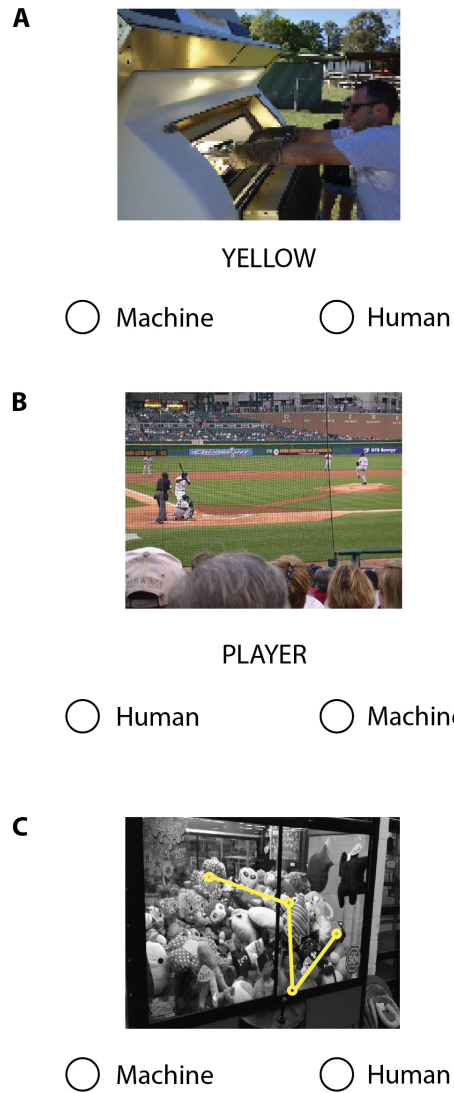


Figure S13: **Schematic illustration of Turing tests for three vision tasks.** See also **Fig. 1** for Turing tests for the three language tasks. **A. Color estimation.** Participants were presented with an image and a color and had to indicate whether the color selection was made by a human or a machine. **B. Object detection.** Participants were presented with an image and a noun and had to indicate whether the object description was made a human or a machine. **C. Attention prediction.** Participants were presented with an image and a sequence of positions (yellow circles) joined by lines and had to indicate whether those locations were the product of human eye movements or machine attention predictions.



Figure S14: **Full results of the Turing test for each Vision task.** Turing test results for human judges (left, **A**, **C**, **E**) and AI judges (right, **B**, **D**, **F**) in the three Vision tasks: Color detection (**A**,**B**), Object detection (**C**,**D**), and Attention prediction (**E**,**F**). The full confusion matrices follow the same conventions as **Fig. 3**. See the color bar in **B** which applies to all panels. The boxes with a black frame denote the best algorithm in terms of its ability to pass as human, i.e., highest $p(H|M)$.

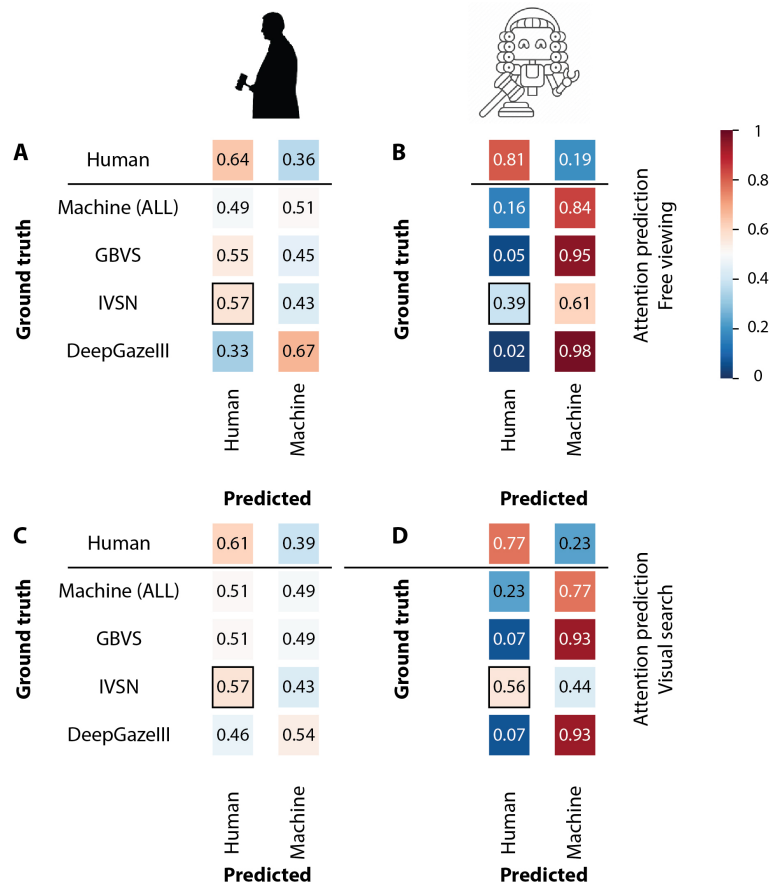


Figure S15: Comparison between free viewing and visual search in the Attention task. Using the same format as in Fig. S14, the results in Fig. S14E are shown here separately for free viewing (A) and visual search (C) and the results in Fig. S14F are shown here separately for free viewing (B) and visual search (D).

Supplementary Tables

Task	Num. Stimuli	Num. Human Agents	Num. Human Judges	Num. Turing Tests	Sources of Datasets	AI agents	AI judges
Image captioning	1,000	229	323	9,400	self-collect, MSCOCO [86], nocaps [81]	GIT-Large [1], OFA-Huge [2], BLIP-Large [3], ClipCap-Transformer (beam search) [4], Microsoft's Azure Cognitive Services [5]	SVM-GPT-curie
Word association	1,500	40	101	2,773	self-collect,	Word2Vec [7], GPT2 [8], GPT3-embedding (davinci) [9], GPT3-prompt (text-curie-001) [9], GPT3-prompt (text-Davinci-002) [9]	SVM-Word2Vec, SVM-GPT2, SVM-GPT3(davinci)
Conversation	2,460	150	367	7,717	self-collect, Topical-Chat [12]	GPT3-text-davinci-002[10], GPT3-text-curie-001[10], Blenderbot[11], DialogPT [12], ChatGPT [13], GPT3.5-turbo-1106 [13]	SVM+BERT [33], ChatGPT-zero-shot, ChatGPT-one-shot
Color estimation	785	45	65	1,625	self-collect, MSCOCO [86]	Google Vision API, Microsoft Azure Cognitive Services [5], MMCQ [91]	SVM+VGG+BERT [33]
Object detection	808	45	79	1,975	self-collect, MSCOCO[86]	Google Vision API, Microsoft Azure Cognitive Services [5], Amazon Rekognition, Detectron2 [15]	SVM+BERT [33]
Attention prediction	547	40	191	2,160	NatureDesign [21], FindingWaldo [21], NatureSaliency [85]	IVSN [21], [22], [85], DeepGaze3 [93], GBVS [23]	SVM on 2D coordinates
Total	7,100	549	1,126	25,650	-	26	10

Table S1: **Specifications of six Turing tasks.** Source datasets, number of Turing tests conducted, number of stimulus, and number of AI models used to collect responses are listed for each task. See **Methods** for task descriptions.