**Treball de Recerca**

**TEST DE TURING**

ICCIC Batxillerats

Ignacio Antón Martín

Lucas Díaz López

1rF

Jordi Borràs

2022/23

**INTRODUCCIÓ:**

**Objectiu del treball:**

El següent treball pretén fer un estudi sobre intel·ligència artificial (IA), centrat en les interaccions d’aquesta amb els éssers humans. L’objectiu principal de l’estudi serà avaluar la capacitat d’un ésser humà per diferenciar entre una IA i un altre ésser humà en sotmetre’l a un exercici.

Obrim la recerca amb les següents preguntes:

*Són els humans capaços de diferenciar, a través d’una conversa, si interactuant amb un altre ésser humà o amb una màquina?*

*És un xatbot entrenat per uns alumnes de batxillerat capaç de superar un test de Turing?*

*Influeix la quantitat d’entrenament del xatbot en la dificultat del subjecte per esbrinar qui és l’ésser humà i quina la màquina?*

En el present treball, definirem *xatbot* com un bot, que mitjançant text, simula mantenir una conversació i entenem com *test de Turing* a l’examen de la capacitat d’una màquina per exhibir un comportament intel·ligent similar al d’un ésser humà.

Per respondre-les dividirem la recerca en dues fases.   
La primera fase serà teòrica i inclourà l’adquisició de coneixements per tal de poder programar un xatbot i entrenar-lo. A la segona fase sotmetrem el xatbot a una prova pràctica per avaluar si és capaç de superar un test de Turing. Amb el nostre *xatbot* intentarem respondre les nostres preguntes d’una manera objectiva mitjançant l’experimentació.

La nostra hipòtesi és que *un xatbot entrenat per nosaltres no serà capaç de superar un test de Turing*, és a dir que no podrà fer creure a una persona que està parlant amb un altre ésser humà.

**Motivació:**

La principal motivació per iniciar i endinsar-nos en aquesta recerca és el fet que a tots dos ens interessa i apassiona el món de la tecnologia i a la vegada tenim inquietuds per l’experimentació.

La nostra necessitat d’aplicar allò que coneixem de forma teòrica és la que ens ha portat a dissenyar un treball amb un alt contingut pràctic.

**Rellevància:**

Pensem que el tema del nostre treball de recerca té una rellevància científica (perquè segueix el mètode científic), tecnològica (perquè aborda una qüestió que està revolucionant el món tecnològic actual) i informàtica (perquè implica programar un xatbot propi)

Però, per una altra banda, també pensem que pot tenir rellevància en un àmbit social i filosòfic, perquè planteja un debat no resolt sobre si les intel·ligències artificials són capaces o no de substituir als éssers humans en qualsevol àmbit.

**ESTAT DE LA QÜESTIÓ:**

En l’actualitat la intel·ligència artificial és una tecnologia molt estudiada i que presenta molta més àrea d’estudi, per això té tantes aplicacions, planejades o que ja estan sent utilitzades. Un dels usos que té més atracció pública són els assistents virtuals, ja que permeten agilitzar i facilitar tasques diàries i tenen un mercat en creixement. També hi és presents en els anuncis, en fer servir xarxes socials, els anuncis que surten a pantalla estan basats en un perfil que ha estat creat sobre nosaltres per una intel·ligència artificial, basat en el que busquem, el que ens agrada i del que parlem.[[1]](#footnote-2)

Des del naixement del test de Turing les màquines han evolucionat fins a punts que ni tan sols Alan Turing va preveure. Això ha provocat que el test que inicialment es va proposar no sigui tant útil com va ser al moment de la seva creació. Ja que s’han desenvolupat mètodes per els que una màquina és casi indistingible d’un humà en els aspectes que Turing qüestiona, però en altres camps les màquines segueixen sent inferiors als humans.[[2]](#footnote-3)

Tot i això, cada any s’estudien a totes les IA disponibles en el mercat i es fa un estudi respecte als resultats, el *ChatBot Report[[3]](#footnote-4).* En aquest estudi s’analitzen les dades, comparades entre IAs i entre anys passats i es fan prediccions respecte als resultats de l’any vinent. S’analitzen matèries com la inversió en IAs, quines regions presenten més evolucions, nous mètodes de processament de dades, la demanda per IAs, l’etapa de les IAs (Inici, grans expectatives, depressió, assimilació i recta de productivitat).

Recentment, l’11 de juny de 2020, va sortir la beta de GPT-3[[4]](#footnote-5), un nou model de llenguatge que s’utilitza per a generar text que sembla humà, el que també ha provocat un gran impuls en el món dels xatbots, ja que permet reconèixer dades i valorar les similituds entre textos escrits per humans.

També s’ha donat, recentment, que un exempleat de Google ha expressat com creu que la IA més avançada de Google (LaMDA), té consciència pròpia, que no hi ha res que la separi d’una ment humana.[[5]](#footnote-6) Tot i que Google ha desmentit aquesta pretensió, dient que simplement l’han entrenat amb bilions de converses, i que només imita molt bé la parla humana, però que no l'entén.

**MARC TEÒRIC:**

Per entendre el funcionament del nostre treball, començarem definint alguns conceptes:

* **IA**

Actualment, segons François Chollet[[6]](#footnote-7), el concepte d’intel·ligència artificial és atribuït a la capacitat d'un sistema per adaptar-se i improvisar en un nou entorn, per generalitzar el seu coneixement i aplicar-lo a escenaris desconeguts.[[7]](#footnote-8)Els sistemes d’ IA solen demostrar alguns dels comportaments associats a la intel·ligència humana com: planificació, aprenentatge, raonament, resolució de problemes, representació del coneixement, percepció, moviment i manipulació i, en menor mesura, intel·ligència social i creativitat.⁴

* **Xarxes neuronals**

Les xarxes neuronals són un mitjà de fer *machine learning*, en què un ordinador aprèn a dur a terme alguna tasca analitzant exemples d’entrenament. Per exemple, un sistema de reconeixement d'objectes podria ser entrenat amb milers d'imatges etiquetades de cotxes, cases, etc., i trobaria patrons visuals en les imatges que es correlacionen sistemàticament

amb etiquetes particulars.  
  
Una xarxa neuronal consisteix en milers o fins i tot milions de nodes de processament simples que estan densament interconnectats. Un node individual hauria d'estar connectat a nodes d’una capa inferior, per on rep la informació, i a nodes d’una capa superior, cap on envia la informació. Seguint aquest sistema de capes la informació passa de tal manera que es mou a través d'elles en només una direcció.

Per cada connexió de nodes s’atribueix un valor anomenat ‘pes’. Quan la xarxa està en funcionament, el node en qüestió rep un valor d’informació el qual es multiplica pel ‘pes’. El producte final es compara amb un valor llindar, si el producte és inferior a aquest el node no passarà informació, si és superior el node enviarà el producte a capes superiors.

Quan una xarxa neuronal està en procés d’entrenament, els ‘pesos’ i els valors llindars inicialment són aleatoris. Entrenant s’alimenta amb informació la capa més baixa de la xarxa (capa d’entrada), i la informació sortirà per l’última capa ( capa de sortida), radicalment canviada. Constantment, s’ajusten els ‘pesos’ i els valors llindars fins que els resultats obtinguts siguin semblants.**[[8]](#footnote-9)**

* **Tensorflow**

Tensorflow és una plataforma de codi obert que ofereix biblioteques i recursos de l'àmbit de la IA. [[9]](#footnote-10)

**Com funciona un xatbot RASA?**

Al nostre experiment farem servir Rasa[[10]](#footnote-11),

Rasa és un servei enfocat a la construcció de xatbots mitjançant python, NLU[[11]](#footnote-12) (natural language understanding) i xarxes neuronals extretes de tensorflow.   
Per entendre el funcionament del nostre xatbot cal entendre alguns principis bàsics que utilitza Rasa:

1. NLU (natural language understanding ):

El funcionament del xatbot està basat en NLU (és el que en IA s'encarrega de la comprensió lectora d'una màquina). L'objectiu del NLU és entendre el sentit comú de les paraules: com aquestes es comporten, les interaccions lingüístiques entre humans, etc. El programa hauria de ser capaç d'entendre intencions, raonament hipotètic, plans i finalitat en un text.

L'enteniment dels textos es fa a través de diferents passos:

* 1. **Tokenització**: Separació del text en tokens.
  2. **Caracterització**: Generació d'informació numèrica pels models de *machine learning*.
  3. **Classificació d'Intencions**: Fent servir els tokens i la informació numèrica s’extreuen les intencions del text.
  4. **Extracció d'entitats**: Extracció de la informació que es vol recordar (noms, correus electrònics, números de telèfon...)[[12]](#footnote-13)

1. Elements bàsics de RASA:

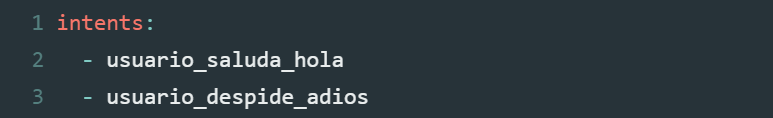
Una vegada conclòs l’enteniment dels textos, el xatbot requereix diferents elements per començar una interacció:

* 1. **Intencions**: Són les intencions dels textos (saludar, acomiadar-se, demanar ajuda…)[[13]](#footnote-14)
  2. **Entitats**: Informació més rellevant del text, la que cal recordar (noms, correus electrònics, números de telèfon…)
  3. **Respostes**: Les respostes que el xatbot donarà depenent de les intencions dels missatges.
  4. **Històries**: Les històries són converses d'exemple que entrenen al xatbot per respondre correctament depenent del que l'usuari hagi dit anteriorment a la conversa.
  5. **Normes**: Les normes descriuen parts de la conversa que sempre han de seguir el mateix camí, independentment del que s’hagi dit anteriorment a la conversa.

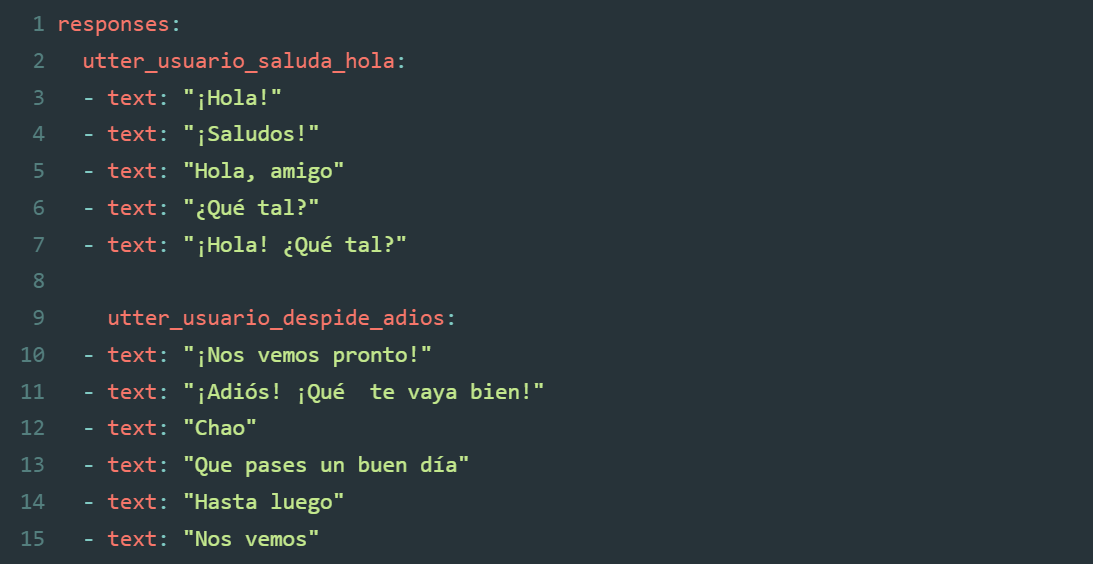
1. Entrenament xatbot RASA:

Dins d’un xatbot de Rasa trobem diferents documents en llenguatge YAML[[14]](#footnote-15) en els quals es troba tota la informació necessària per a l’entrenament del xatbot. Rasa utilitza YAML per gestionar d’una forma ordenada totes les dades d’entrenament. D’aquests documents cal destacar-ne: domain.yml, nlu.yml, rules.yml, stories.yml.

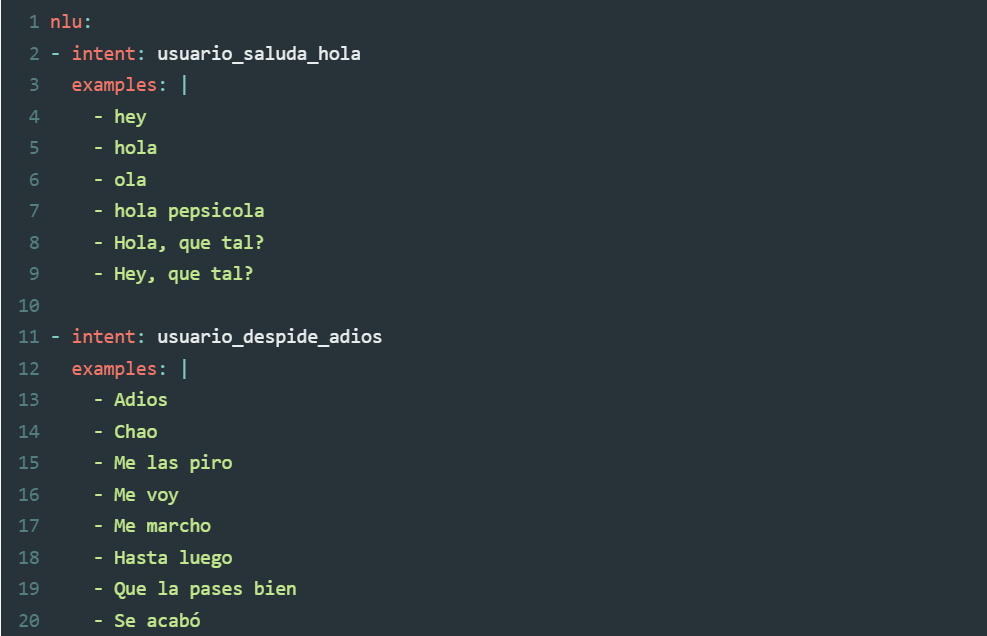
* 1. **Domain.yml**: Aquest document especifica i defineix les intencions, les respostes i les accions que el xatbot hauria de conèixer. És aquí on es creen les intencions necessàries per originar les converses desitjades (com podria ser la intenció de saludar o d’acomiadar-se). Aquestes intencions seran les que el xatbot identificarà en els missatges de l’individu amb el qual parli:[[15]](#footnote-16)



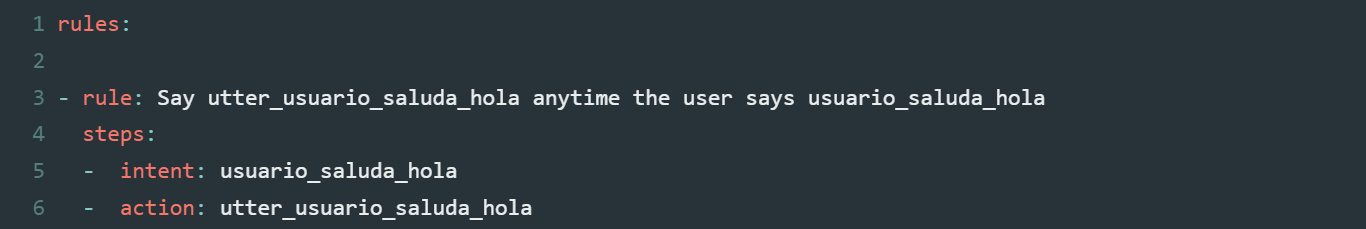
També el *domain* és el lloc on es creen les accions i les respostes del xatbot. Per cada acció cal especificar-hi un nom i afegir-hi una o més respostes textuals. Aquí en trobem un exemple:

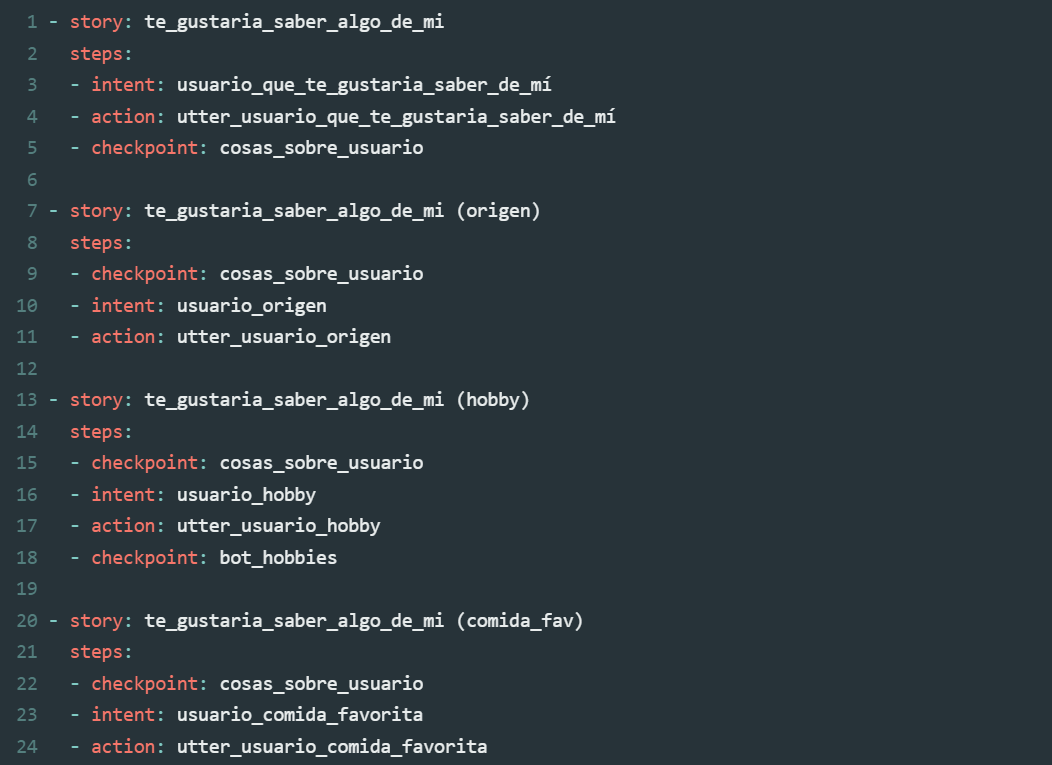
****

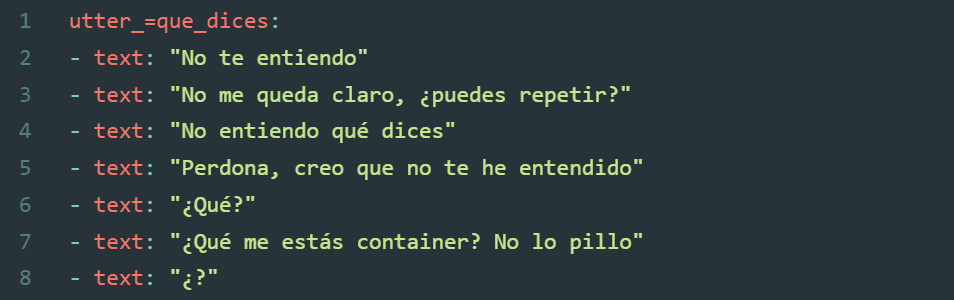
* 1. **nlu.yml**: Aquest document conté exemples dels missatges que el xatbot pot rebre i de la intenció que tenen. Cada intenció creada al *domain* requereix exemples. Amb els missatges d’exemple el xatbot serà capaç d’identificar les intencions dels missatges rebuts mitjançant un procés de NLU (Tokenització, Caracterització, Classificació… ). Com més semblant siguin els missatges rebuts i els missatges d’exemple més fàcil serà la identificació d'intencions[[16]](#footnote-17). També, la major quantitat de missatges d’exemples facilitarà la feina al xatbot



* 1. **rules.yml:** Aquest document s’encarrega de definir les normes del xatbot. Les normes descriuen fragments breus de converses que sempre han de seguir el mateix camí. En aquestes normes s’indica al xatbot quines accions ha de seguir quan identifiqui intencions. .[[17]](#footnote-18)



* 1. **stories.yml**: En aquest document trobem històries que són representacions d’una conversa entre xatbot i usuari, en format d’intencions i respostes. Les històries són converses d’exemple que poden seguir diferents camins sense estar limitades com les normes.Aquest document facilita al xatbot el seguiment de contextos en una conversa.[[18]](#footnote-19)
  2. **FallBack**: En cas que el xatbot no pugui identificar quina intenció té el missatge, activa el FallBack i dona una resposta generalitzada.[[19]](#footnote-20)



**Què és el Test de Turing (TT)?**

El test de Turing[[20]](#footnote-21) és un examen de la capacitat d’una màquina per exhibir un comportament intel·ligent similar al d’un ésser humà. L’objectiu de la màquina és fer creure a un jutge humà que ella és un altre humà.

El nostre experiment es basarà en el model inicial de TT que va proposar Alan Turing després d’haver adaptat informàticament el joc ’*Imitation Game[[21]](#footnote-22).*

En la versió original de l’*Imitation Game*  hi ha tres subjectes:

* una dona (*A*)
* un home (*B*)
* i un jutge(*C*).

La dinàmica del joc és la següent: el jutge (C) es comunica amb la dona (A) i amb l’home (B) a través de preguntes amb l’objectiu d’identificar quin dels subjectes és la dona i quin l’home.

El jutge pot fer tota mena de preguntes i A i B hauran de respondre sempre. L’home i la dona poden adoptar diferents estratègies per evitar ser descoberts pel jutge (per exemple mentir, incriminar a l’altre, respondre amb evasives, etc.)

Alan Turing utilitzà aquest joc i el va adaptar proposant canviar a un dels subjectes (A o B) per una màquina. El jutge(C) s’hauria de comunicar a través d’un xat digital amb l’objectiu de diferenciar l’ésser l’humà de la màquina.

Perquè aquest joc sigui vàlid caldrà reduir al màxim la diferència entre ésser humà i màquina. A més, el jutge no podrà veure físicament ni a l’ésser humà ni a la màquina i només podrà fer servir el xat digital per comunicar-se.

**METODOLOGIA:**

El nostre treball de camp consistirà en el disseny i posada en pràctica d’un TT.

Fase prèvia:

Es programarà un xatbot capaç de mantenir converses senzilles i de respondre a preguntes superficials (edat, nom, origen, residencia…). També se'l programarà amb la intenció de simular un comportament humà.

Test de Turing:

Participants:

* **Jutge/subjecte** (individus entre 13 i 19 anys)
* **Màquina (xatbot)**: ordinador
* **Persona/investigador**: Ignacio
* **Controlador extern**: Lucas

Mostra:

Es calcula una mostra de 27 participants.

Desenvolupament del TT:

El **jutge** tindrà dues converses per xat. Una de les converses serà amb la **persona** i l’altra amb el **xatbot**. El subjecte coneixerà prèviament l’existència del xatbot.   
  
L’objectiu del **subjecte** serà identificar qui és el xatbot i qui la persona.   
L’objectiu del **xatbot** serà convèncer al subjecte de que ell és la persona.

L’objectiu de l’**investigador** serà convèncer al subjecte de que ell és la persona.

Tant **xatbot** com **investigador** hauran d’aconseguir simular un comportament humà.

Cadascun dels subjectes haurà de fer 2 tests amb el xatbot. Cada test tindrà un nivell d’entrenament diferent basant-nos en una hipòtesi prèvia segons la qual el grau d’entrenament pot influir en la capacitat de resposta del subjecte i en el seu grau d’encert.

Es dissenyaran 2 nivells de dificultat en el xatbot (en funció de la capacitat prèvia d’entrenament) i s’aplicaran tots dos nivells a cada subjecte de forma consecutiva.

Un cop realitzats ambdós tests, s’obtindrà dades de cadascun dels subjectes relacionades amb els nivells de dificultat i els temps de resposta que han necessitat per resoldre la prova.

Els subjectes només podran fer servir un màxim de missatges i l’idioma utilitzat serà el castellà. Només sabran els seus resultats (si han encertat o no) una vegada acabades les dues proves.

Variables:

Es mesuraran les següents variables:

* + **Temps**: El temps que els subjectes trigaran a identificar el xatbot per saber si existeix relació amb el nivell d’entrenament del xatbot. Suposem que la relació serà lineal: a major entrenament del xatbot, més temps es requerirà per identificar-lo.
  + **Nombre de missatges escrits:** Es quantificaran el nombre de missatges escrits i utilitzats fins que el subjecte finalitzi el TT. Suposem que existirà una relació entre el nombre de missatges i el grau d’encert del subjecte. També valorarem si existeix relació combinada entre el nombre de missatges, el temps i el grau d’entrenament.
  + **Grau d’encert** dels subjectes: Com a mesura obligatòria, s’observarà si els subjectes encerten (és a dir si identifiquen correctament el xatbot i la persona). Creiem que aquesta mesura és la més important i la que serà capaç de respondre les nostres preguntes de recerca i a sustentar o no, la nostra hipòtesi inicial.
  + **Nombre de caràcters i paraules**: Es mesurarà la quantitat de caràcters i paraules per missatge per valorar si existeix relació amb el nivell d’entrenament i el grau d’encert tot i que sospitem que aquestes mesures podrien ser més secundàries i probablement no hi hagi una relació directa.
  + **Anàlisi qualitativa**: Per últim, s’estudiarà, de forma qualitativa, el tipus de converses realitzades en els tests. També s’analitzarà les característiques dels missatges determinants (definirem com missatge determinant a aquells missatges que han facilitat que els subjectes es decideixin en la seva resposta final). En aquest apartat, pretenem trobar les falles del xatbot, si aquest és capaç de seguir contextos, si pot respondre preguntes, etc. A dins d’aquesta anàlisi, s’inclourà una pregunta final en la qual es preguntarà als subjectes perquè han decidit qui és qui i quins són els arguments que han provocat les seves respostes.

Característiques de l’estudi:

**Edat dels subjectes**: L’edat dels subjectes oscil·larà dels 13 als 19 anys. El fet d’escollir aquest rang d’edat és rellevant, ja que considerem que si es duguessin a terme els tests a gent de qualsevol edat, trobaríem resultats totalment diferents. A causa que la majoria d’adolescents conviuen amb aparells electrònics (mòbils, tauletes, ordinadors…) i saben moure's en aquests, considerem que són subjectes exigents. Aquesta exigència és la que busquem en els tests. A més, podrem observar si l’edat (dintre del rang de 13-19) té alguna relació amb la resposta dels tests, però no prendrem l’edat com una mesura de l’estudi, ja que no farem distincions de grups d’edat.

**Llibertat d’acció del subjecte:** Els subjectes podran utilitzar qualsevol estratègia que considerin necessària per a resoldre la prova, com podrien ser: fer preguntes, repetir-se, escriure caràcters aleatoris, escriure els mateixos missatges en totes dues converses, etc.

**Coneixements previs sobre la prova**: Els subjectes seran coneixedors de l’existència del xatbot.

**Xatbot**: El xatbot sempre serà el mateix, però en diferents graus d’entrenament depenent de la prova. Això vol dir que una vegada dut a terme el primer test el xatbot serà intocable, no es podrà entrenar més, ha de quedar igual per tots els subjectes.

**Ésser humà**: La persona amb qui els subjectes conversen sempre serà el mateix integrat del grup. La persona i el xatbot sempre seran els mateixos.

**Màxim de missatges**: Hi haurà un màxim de missatges per cada conversa, el màxim serà de 20 missatges. Si es fa ús de tots els missatges en les dues converses s’hauran utilitzat 40. Cal destacar que quan el subjecte tingui clara la seva resposta ha de contestar, sense necessitat d’utilitzar tots els missatges possibles.

**Via de comunicació (Discord)**: Les converses es duran a terme a Discord. Discord és un servei de missatgeria instantània freeware de xat de veu, vídeo i xat per text. Hem decidit utilitzar aquesta aplicació principalment pel seu sistema de servidors (grup de xats). Dins dels servidors podem crear una gran quantitat de xats ordenats, on es duran a terme els tests. Amb aquest sistema podem emmagatzemar i organitzar les converses d'una manera fàcil i visual. A més a més, amb Discord els tests es realitzaran amb l'anonimat que es busca en *l'Imitation Game*, ja que els usuaris del xatbot i la persona seran els mateixos. La persona i el xatbot comparteixen compte de Discord.

**Mètode de respostes**: Per últim, el xatbot i la persona tindran les mateixes possibilitats a l’hora de respondre, és a dir, la persona només podrà conversar amb les respostes/missatges que té el xatbot. La persona tindrà dues llistes amb totes les respostes possibles del xatbot. La primera llista contindrà les respostes/missatges que té el xatbot menys entrenat (Fase 1), i la segona tindrà totes les respostes/missatges del xatbot més entrenat (Fase 2).

Creiem que aquesta limitació és rellevant i necessària. Considerem que si la persona pogués conversar amb llibertat, sense necessitat d’usar les respostes del xatbot, els tests serien resolts molt senzillament. Pressuposem que una persona té un nivell de conversa molt més elevat que el del xatbot que entrenarem. Ens trobem davant d'una manca de recursos, de coneixement sobre la matèria i de temps per poder equilibrar aquesta diferència. Per tant, per ser capaços de realitzar aquest estudi assumirem aquesta limitació, per poder observar com dintre d’un rang limitat de respostes/missatges, la persona i el robot triaran.

**Logística per a realitzar els tests:**

Els tests es realitzaran a casa dels membres del grup de recerca (Ignacio i Lucas). Es requeriran tres espais/habitacions diferents per evitar tota comunicació visual entre els investigadors i entre els investigadors amb el subjecte. Si hi hagués més d’un subjecte, es facilitarà un espai d’espera separat (en aquest espai es proporcionaran elements d’entreteniment com podrien ser cartes, jocs de taula… per afavorir la tranquil·litat i col·laboració dels participants) Es farà signar un consentiment informat als participants dels tests.

A cada subjecte participant de l’estudi se li assignarà un nombre cardinal entre l’1 i el 27.

El subjecte estarà situat a l’habitació **número 1 (foto 1)**. A l’habitació 1 hi haurà un ordinador de sobretaula amb dos monitors. Els monitors seran diferenciats amb etiquetes (un serà el monitor A i l’altre el B). També hi haurà un cronòmetre que s’utilitzarà per mesurar el temps trigat en cada test.

Fotografia 1



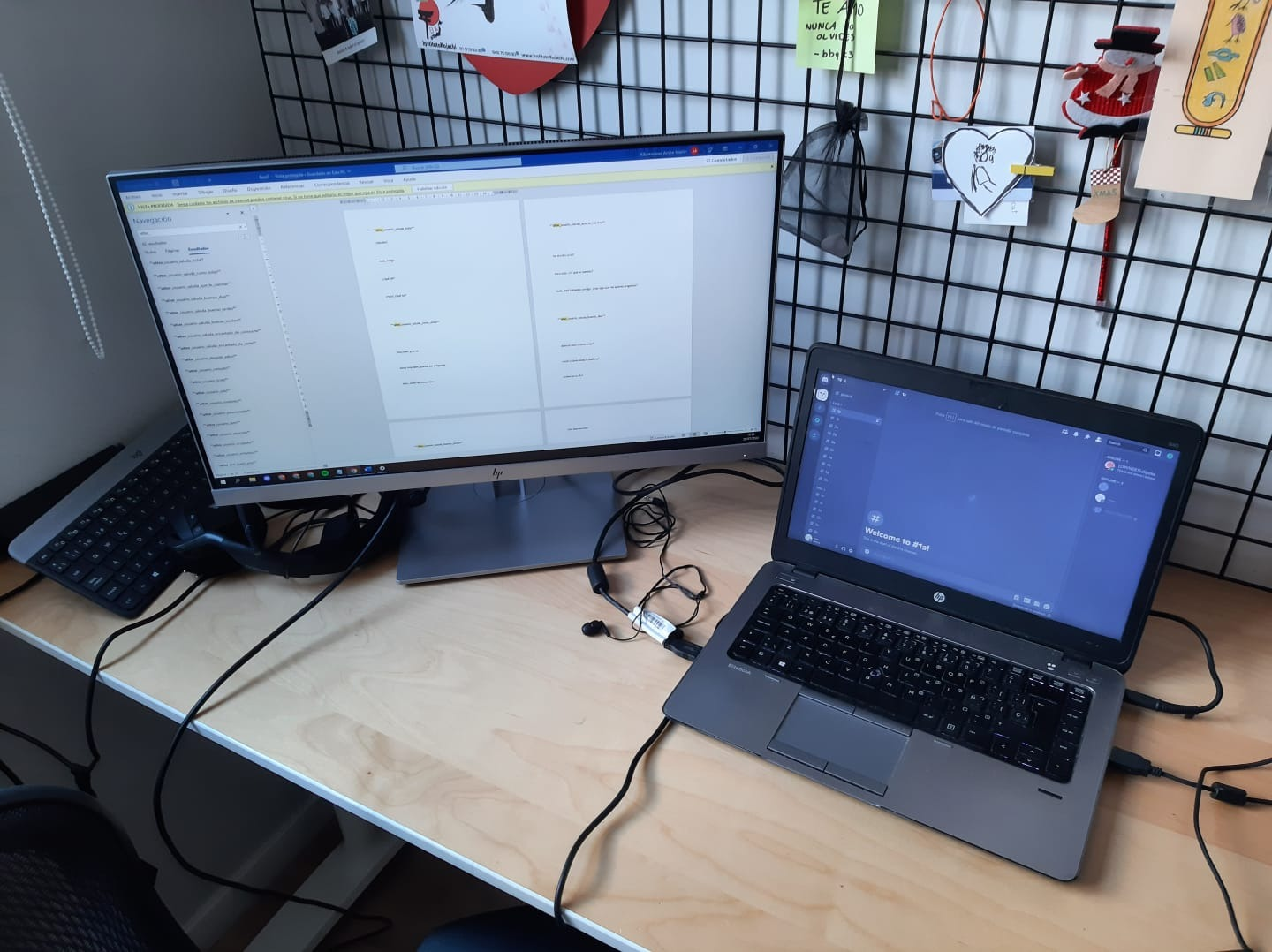
La pantalla de tots dos monitors mostrarà un servidor de Discord amb dos xats totalment iguals, excepte el seu nom. Els xats s’identificaran amb un número i una lletra: el número serà el d’identificació assignat a cada subjecte i la lletra serà la A o la B depenent del monitor (per exemple, en el cas del tercer participant els xats seran el 3a i el 3b).

Conversant darrere dels dos xats hi haurà sempre el mateix investigador (Ignacio) i el xatbot. La decisió de per quin xat parla el xatbot i per quin la persona es farà aleatòriament i amb anterioritat als tests.

Acompanyant al subjecte hi haurà l’altre investigador (Lucas) què haurà de comptar el nombre de missatges enviats per evitar que no se sobrepassi el límit de missatges establerts (20 per conversa) i també cronometrarà el test. Aquest investigador no tindrà cap interacció amb la persona excepte:

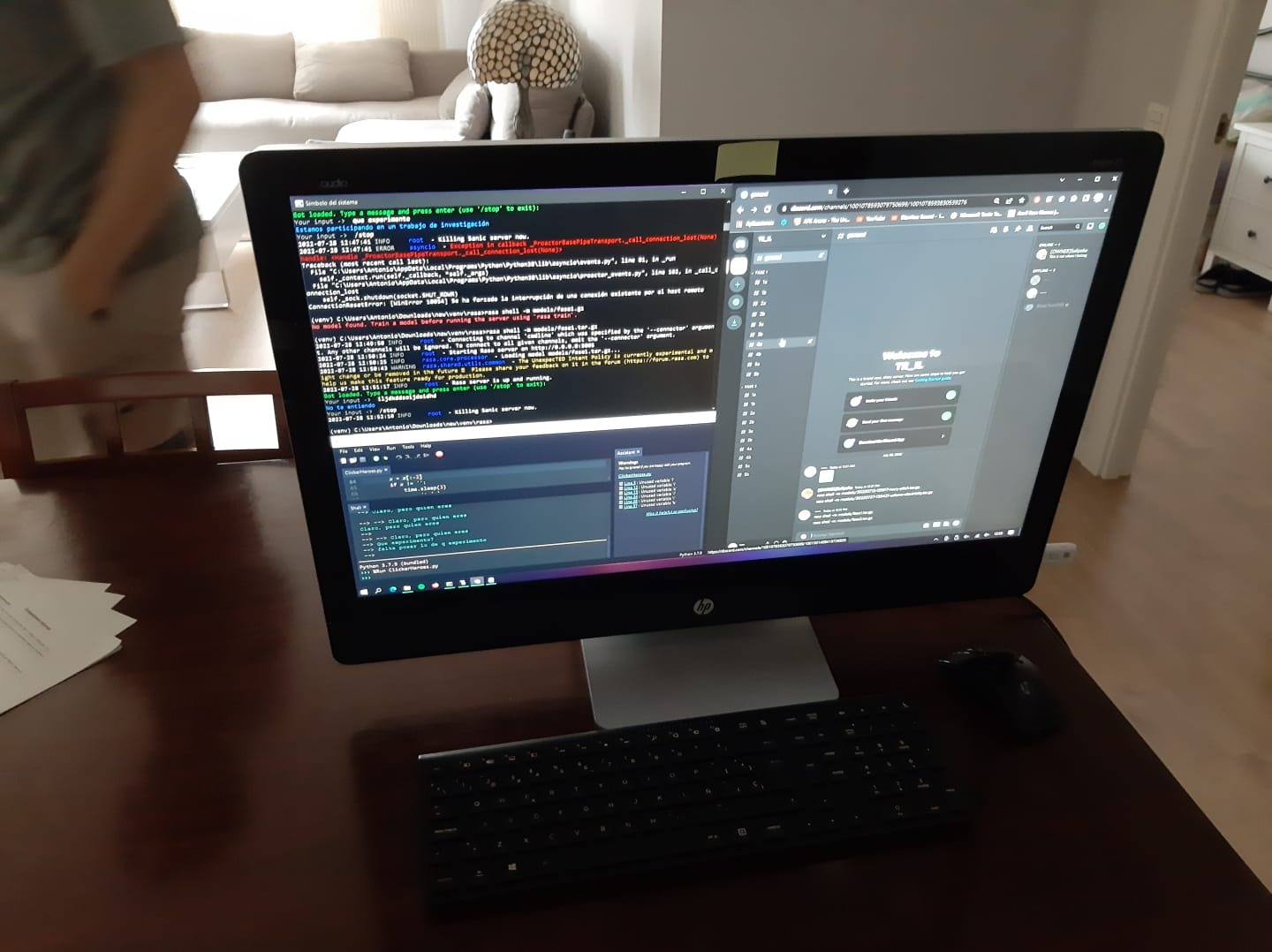
* avisar quan el test comença
* resoldre qualsevol dubte respecte al funcionament del test
* ajudar amb qualsevol problema que pugui sorgir

A l’habitació **número 2** es trobarà el membre del grup que adoptarà el rol de **persona** en el test i hi haurà un segon ordinador. Aquest tindrà la pantalla dividida, a un costat hi haurà el Word corresponent amb les respostes possibles del xatbot i en l’altre costat el xat de Discord. Aquest membre no tindrà cap interacció amb els participants durant els tests. Quan comencin les converses no podrà revisar el que respongui el xatbot per evitar deixar-se influir per ell.



A l'última habitació, la **número 3**, trobarem un últim ordinador.

Hi haurà una persona externa al test. Aquest individu estarà monitorant el **xatbot**, és a dir, copiant les respostes del participant, transferint-les al xatbot i tornant la resposta que el xatbot dona al participant. La persona encarregada d’aquesta tasca no interactuarà amb res més, només és un intermediari entre el xatbot i el participant i no formarà part de l’equip d’investigadors (aquesta tasca és necessària perquè el xatbot conversa al CMD i cal transferir els missatges als xats de Discord)



Una vegada acabats els tests de la fase 1:

Investigador 1 (Lucas):

L’investigador 1 que es troba a l’habitació 1 canviarà els xats als de la fase 2.

Aquest investigador preguntarà al subjecte que ha finalitzat la fase 1 les raons per les quals ha optat per una resposta en concret.

Després marxarà a l’habitació 3 on canviarà el nivell d’entrenament del xatbot.

Investigador 2 (Ignacio):

A l’habitació 2 l’altre membre investigador canviarà el Word amb les respostes del xatbot més entrenat per poder iniciar la fase 2.

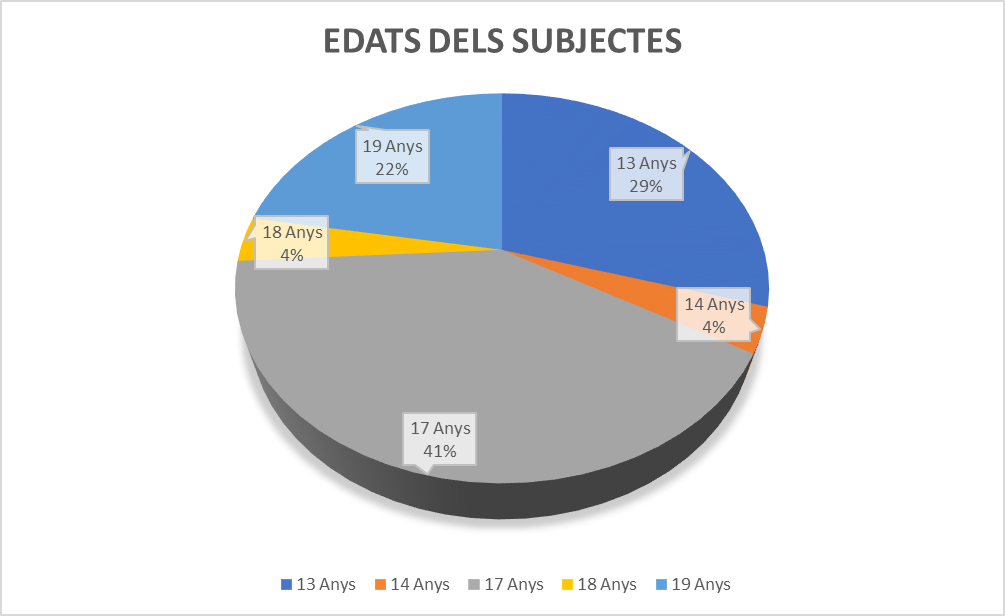
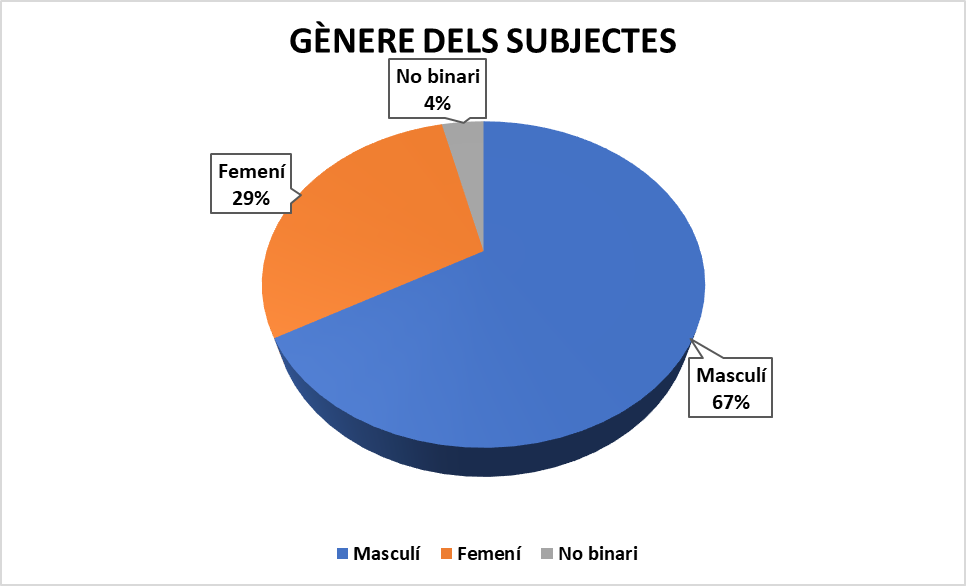
La fase 2 seguirà la mateixa seqüència que la fase 1.

**RESULTATS:**

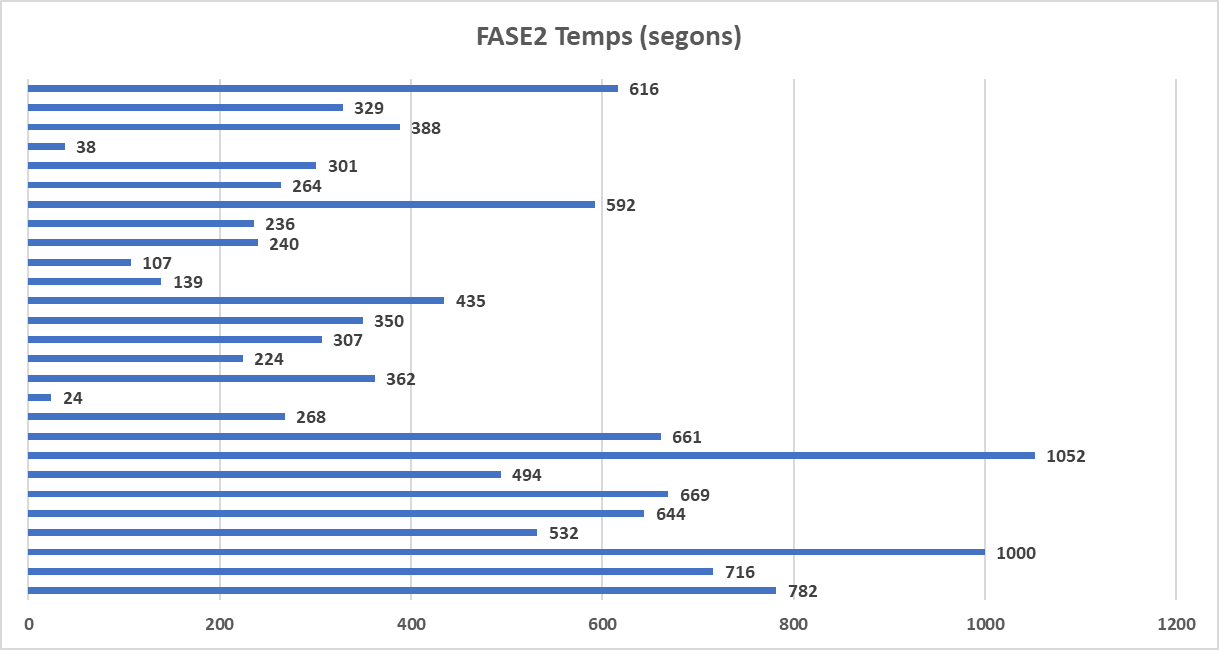
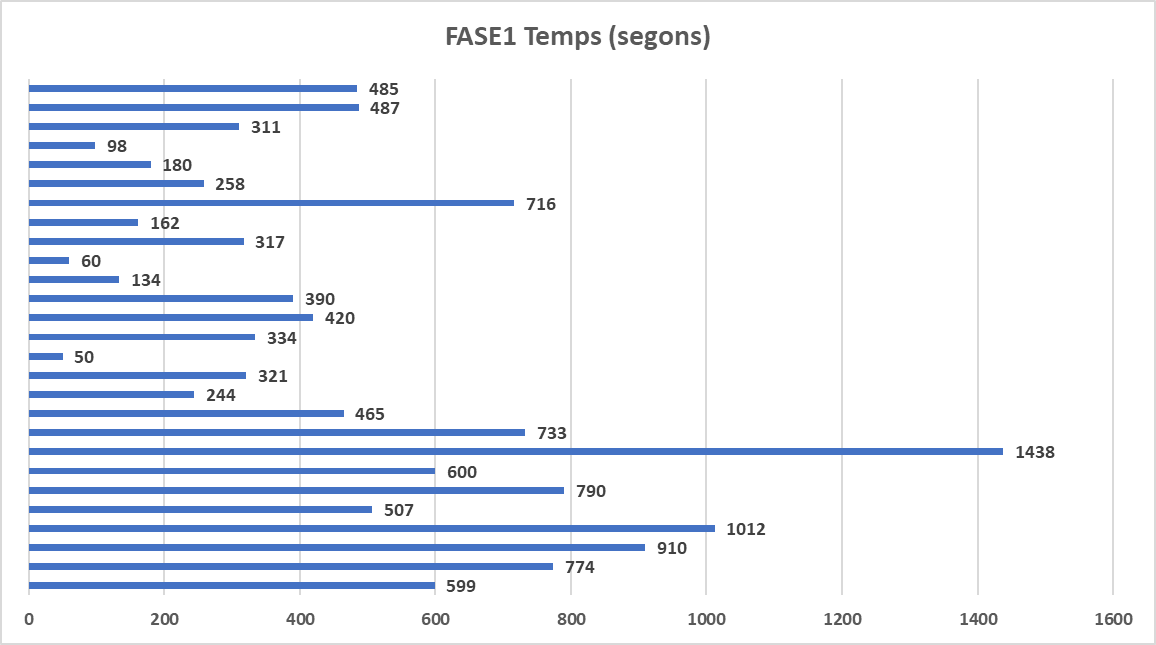
A l’estudi finalment hi ha hagut 27 participants, cadascun dels quals ha realitzat 2 proves.

S’han obtingut 54 converses que posteriorment hem analitzat.:

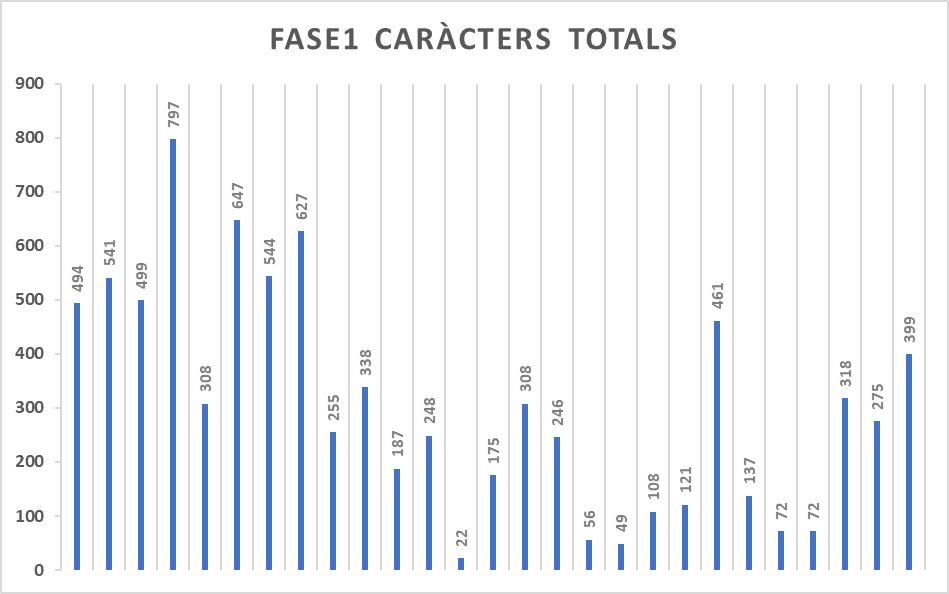
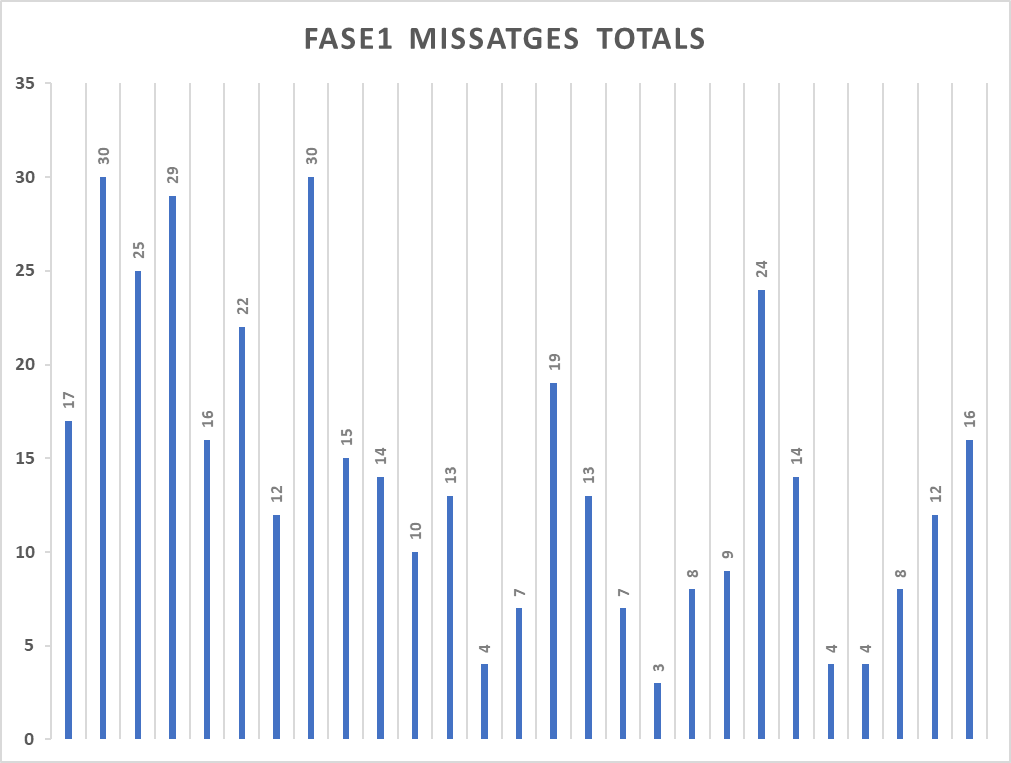
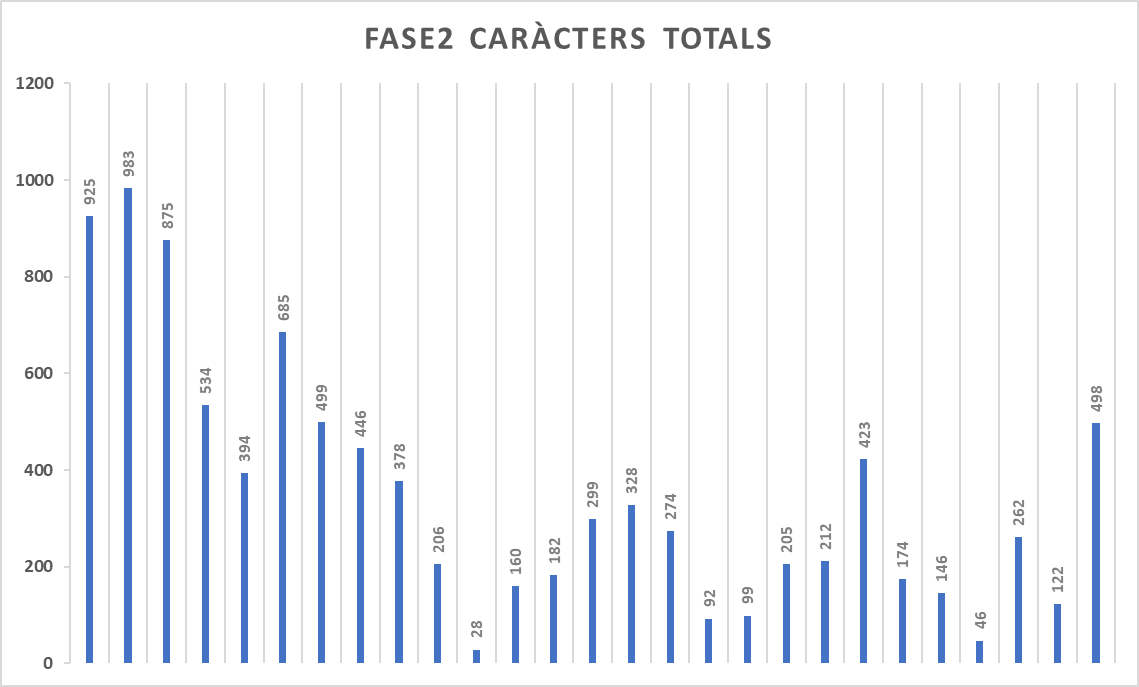
L’edat i gènere dels participants, per tal d’establir la composició de la mostra.



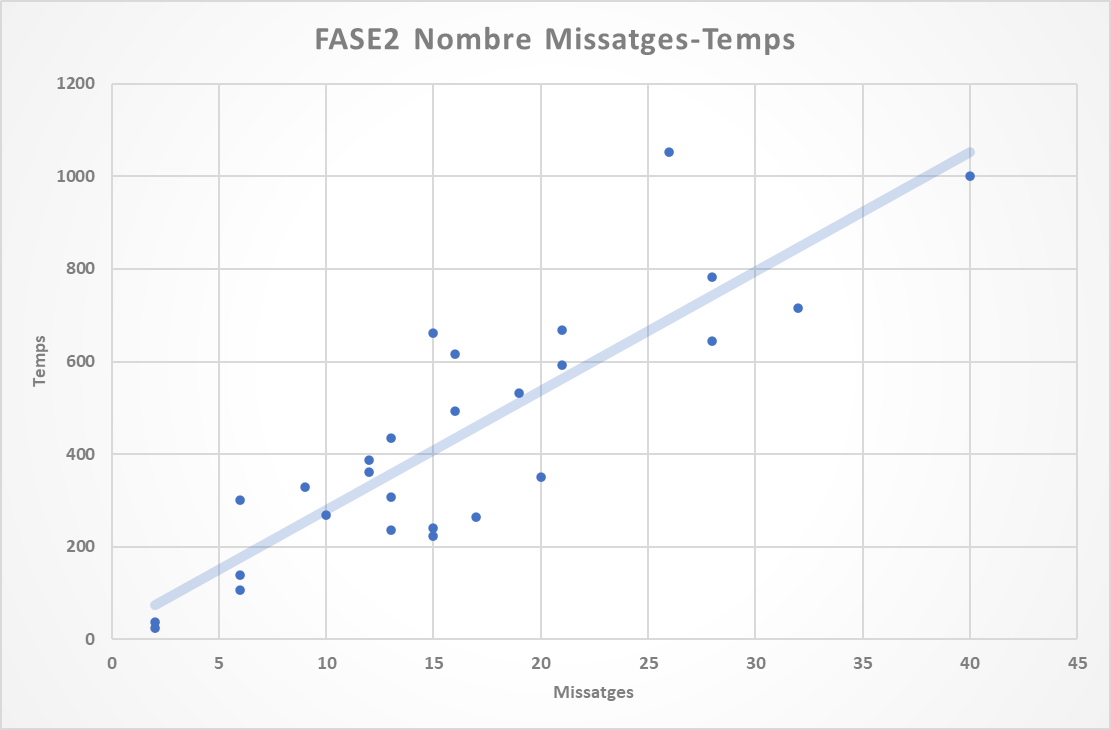
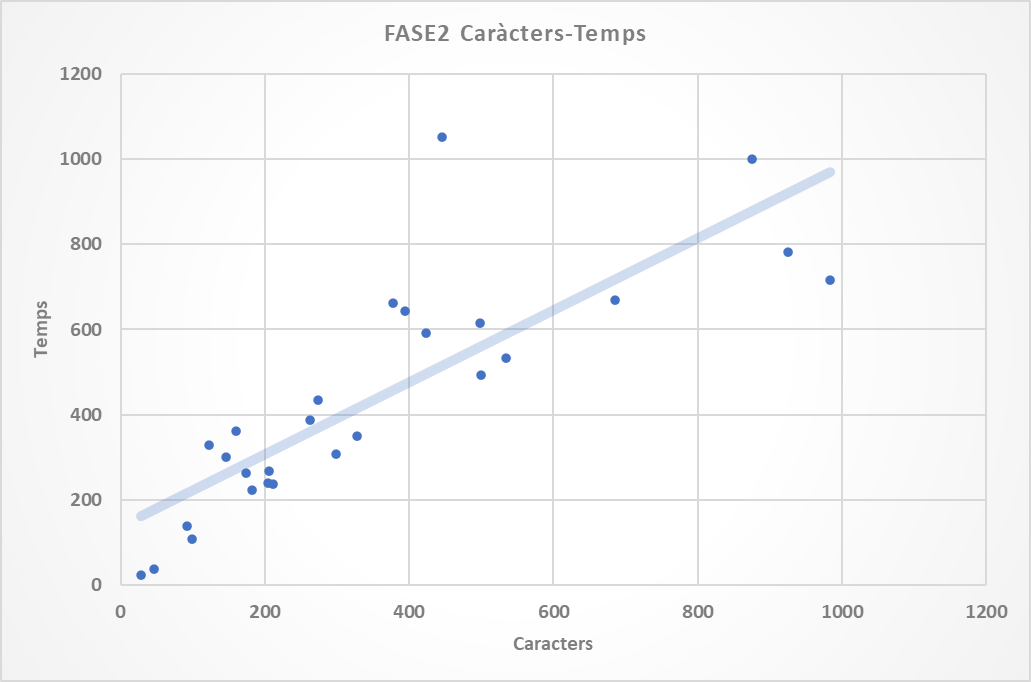
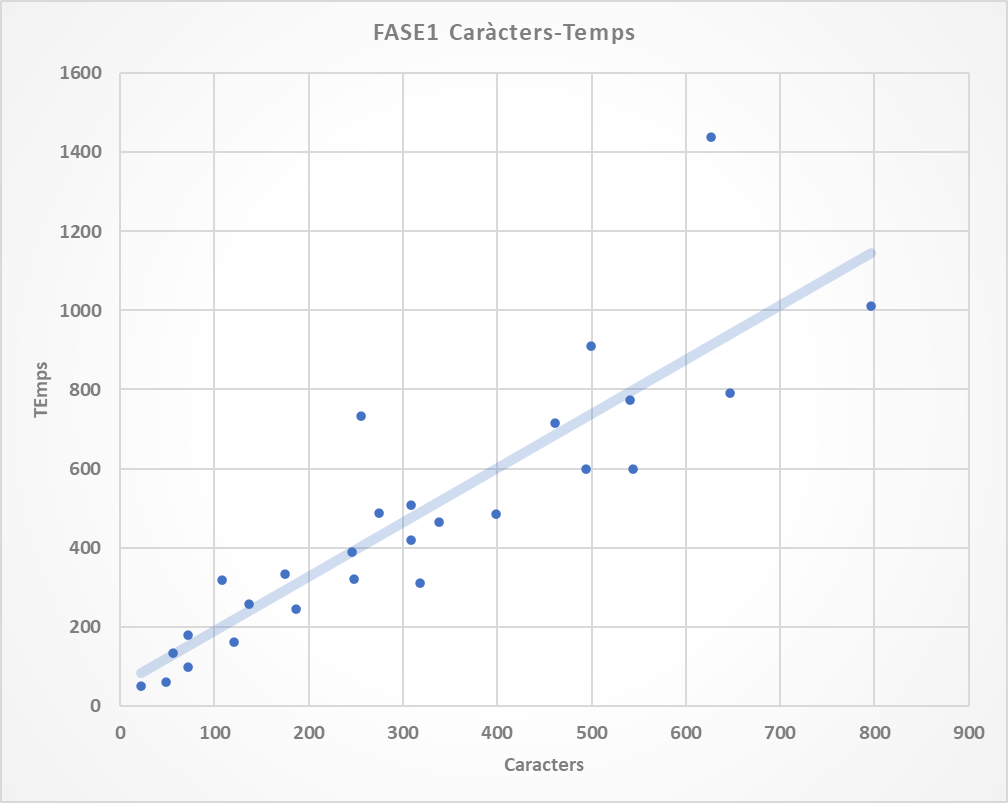
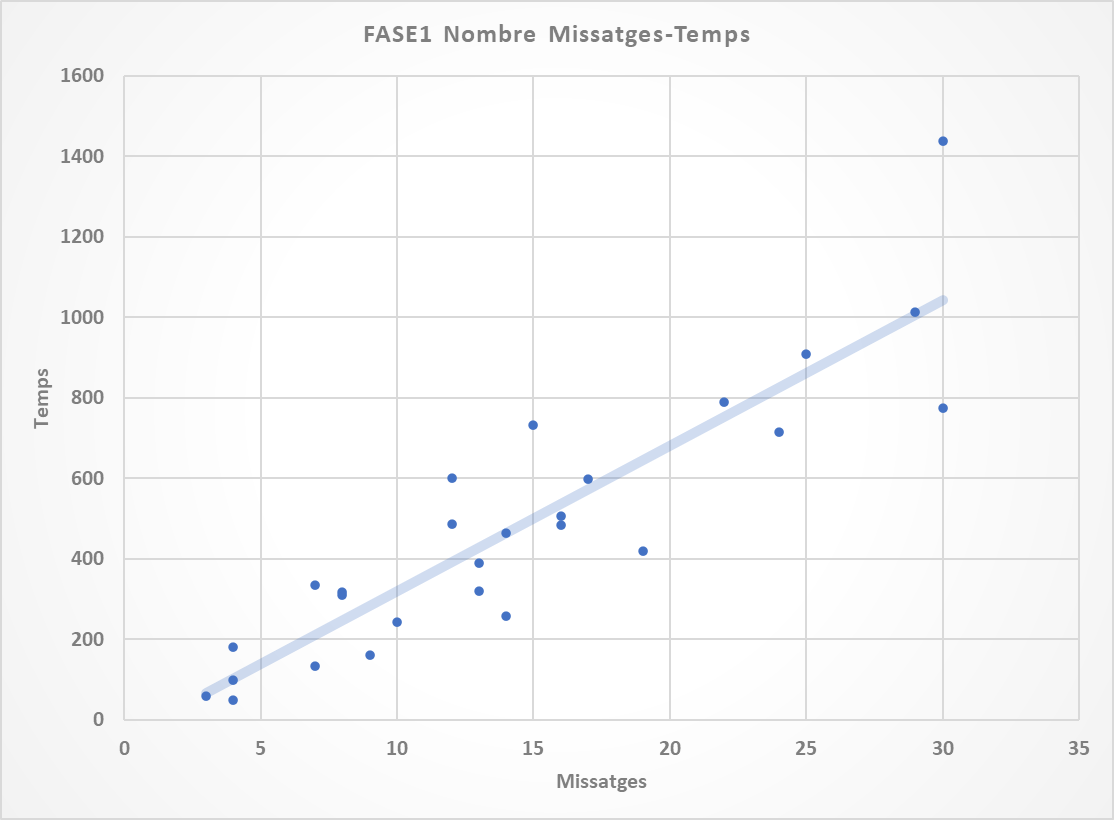
El temps que van trigar els participants (mesurat en segons) en resoldre el temps està representat al gràfic 1. Es van analitzar les dues fases (fase 1 i fase 2) per separat per posteriorment poder-les comparar.

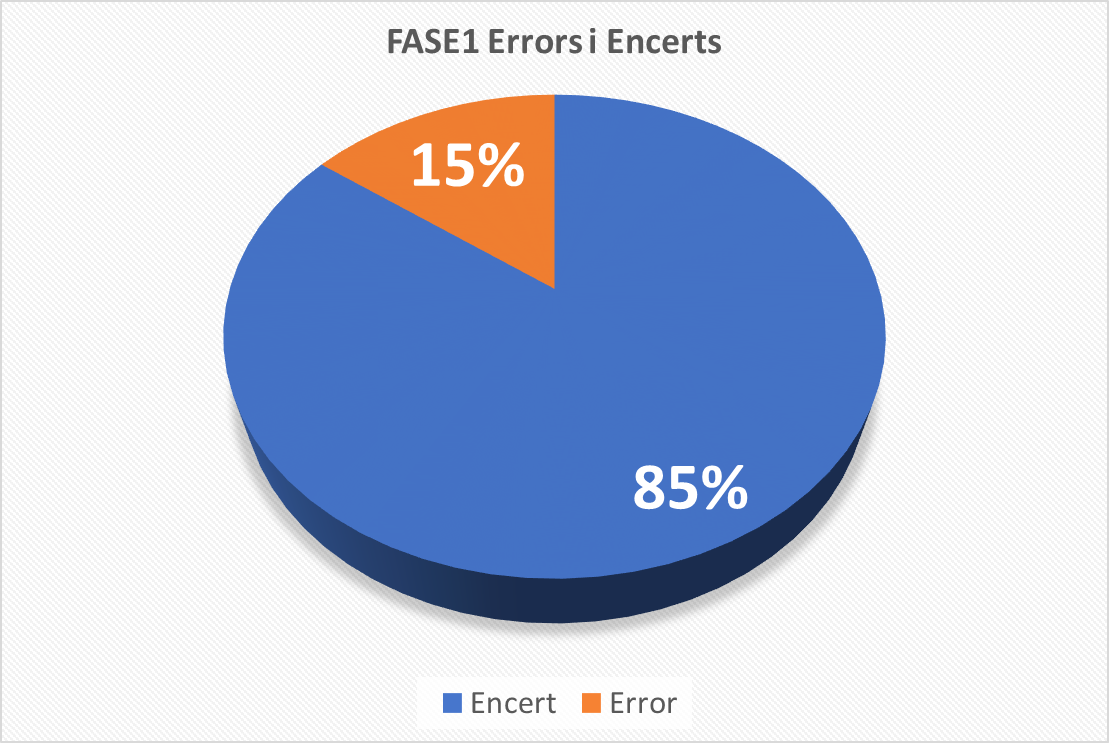


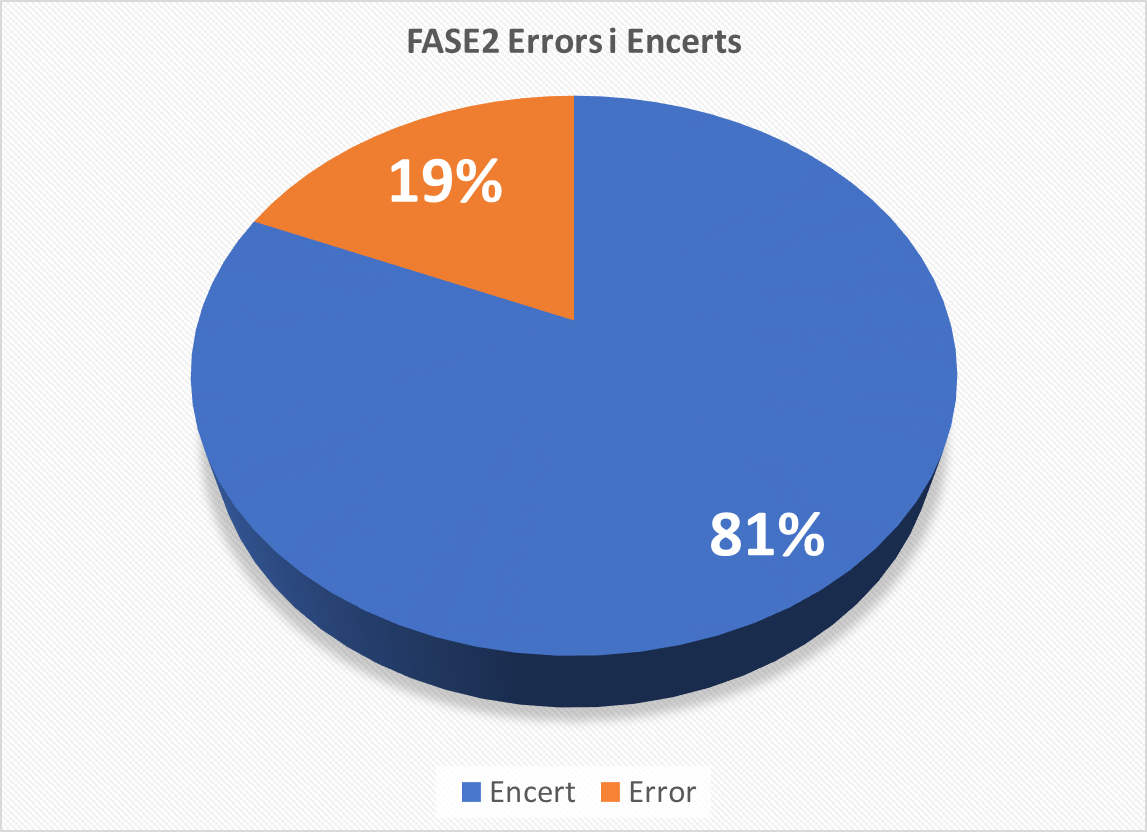
També es van analitzar el nombre de missatges i caràcters emprats per cada subjecte (gràfiques 3 i 4)



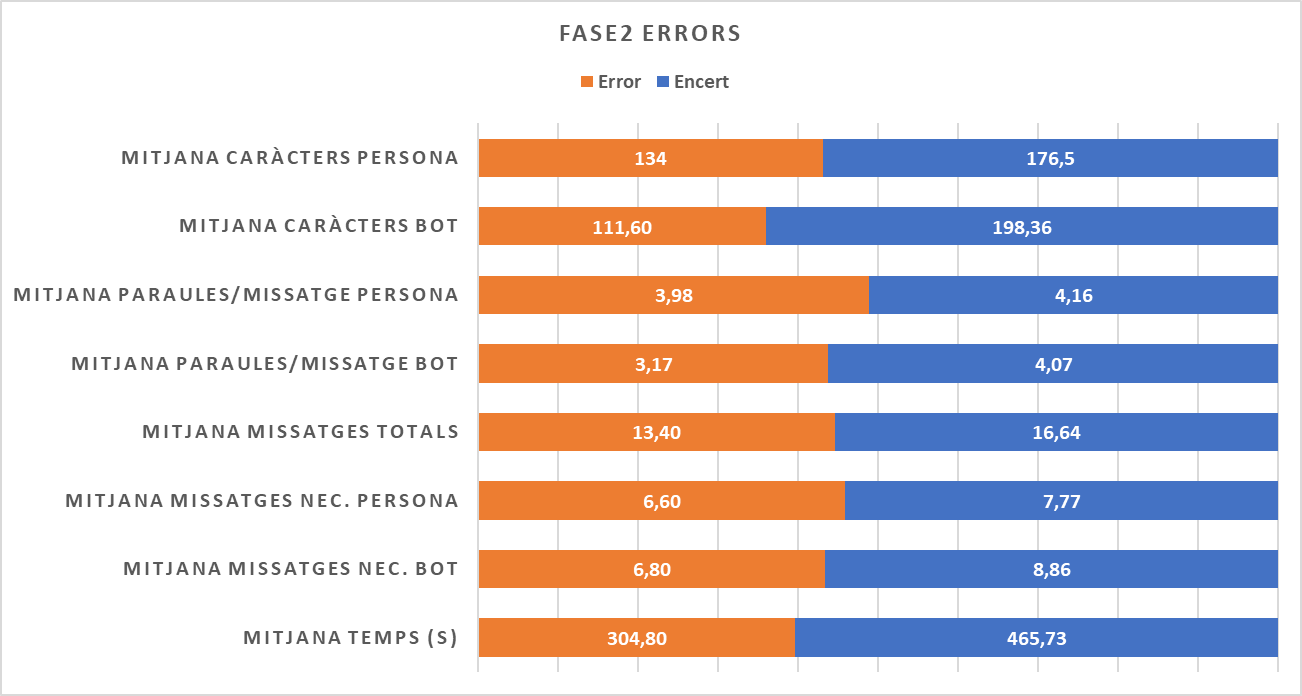
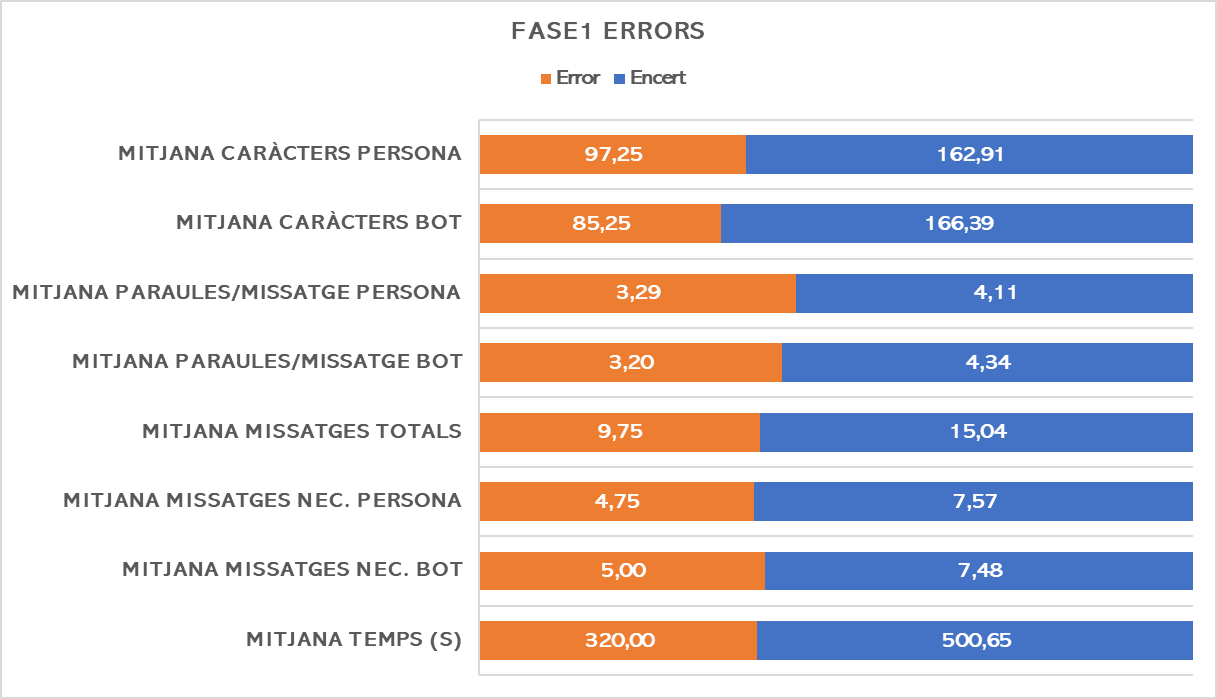
Després es va buscar la relació entre les 3 variables: nombre de missatges, nombre de caràcters i temps.



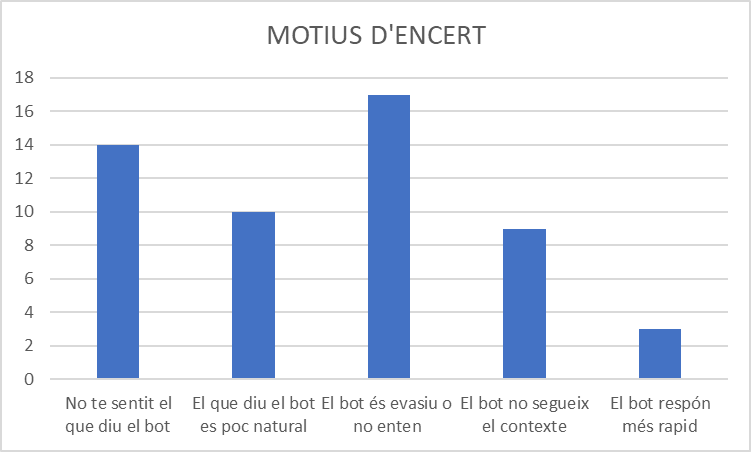
En relació al nombre d’errors (es defineix com error la manca d’encert en identificar xatbot com a xatbot i persona com persona) es van obtenir els següents resultats (figura \_ i \_): 



Es van comparar les diferents variables (nombre de caràcters, paraules, missatges totals….) produïdes en les 2 fases entre les persones que havien errat i les que havien encertat (gràfiques \_ \_)



Finalment, es va fer una anàlisi qualitativa de les raons que havien motivat la resolució de les proves en cada fase en els subjectes que van encertar. (gràfica \_\_\_)



**DISCUSSIÓ:**

La investigació realitzada té com a objectiu demostrar la nostra hipòtesi inicial. Consideràvem que el nostre xatbot seria incapaç de superar un TT.

La nostra hipòtesi inicial contemplava el 100% d’encert en tots els subjectes inclosos en l’experiment (és a dir, que el xatbot no seria capaç de simular un comportament humà). En canvi, els resultats observats no han confirmat això: alguns subjectes han fallat.   
El xatbot va superar el test en 4/27 ocasions en la fase 1 i a 5/27 a la fase 2. Aquests resultats semblen confirmar que el xatbot que hem dissenyat està capacitat per superar el nostre TT. Tot i així no podem assegurar que alguns dels errors comesos pels subjectes no siguin casuals o siguin causats per la sort.

Recuperant una altra de les preguntes d'investigació plantejades a l’inici del treball, ens fixem en el grau d’entrenament del xatbot.   
A la fase 1, on el xatbot està menys entrenat perquè té menys dades d’intenció i resposta, el grau d’encert dels subjectes és més alt (23/27) en canvi, a la fase 2, on el xatbot està més entrenat, el grau d’encert dels subjectes és menor (22/27). No podem afirmar amb rotunditat que com més grau d’entrenament tingui el xatbot més difícil serà la seva identificació, perquè la diferència entre els encerts de la fase 1 i els de la fase 2 és insignificant, però sembla intuir-se que el grau d’entrenament del xatbot pot influir en el grau d’encert. Pressuposem a més, que aquesta diferència (23/27 i 22/27) no és major perquè el factor d’aprenentatge dels subjectes entre les fases 1 i 2 afavoreix el grau d’encert. Aquest factor de confusió no el vam tenir en compte a l’hora de dissenyar l’experiment.

Sense haver contemplat inicialment el factor d’aprenentatge de cada subjecte, vam preveure un temps total per fase d’entre 3-4 minuts per solucionar el test de la fase 1, i de 5-6 minuts per solucionar el test de la fase 2.   
Observant els resultats de l’experiment, la mitjana de temps a la fase 1 ha estat de 7:53 min i la de la fase 2 de 7:15 minuts. Els resultats han evidenciat que la nostra previsió era errònia, ja que el temps empleat per solucionar els tests ha estat major del previst, i a més, en general, els de la fase 1 van requerir de més temps que els de la fase 2. Sospitem que la raó tingui a veure amb el factor d’aprenentatge que hem comentat.

També hem trobat una possible relació lineal entre el temps i el nombre de missatges totals en totes dues fases. Com més temps trigui un subjecte a solucionar el test més missatges escriurà aquest. Observem que també hi ha una relació similar amb el temps i els caràcters escrits. En conclusió, sembla que el temps, el nombre de missatges i el nombre de caràcters són directament proporcionals. Els resultats evidencien que en totes dues fases, els subjectes que no encerten triguen menys temps a resoldre els tests, escriuen menys missatges i menys caràcters que els subjectes que encerten. Assumim que, en general, com menys es trigui a realitzar un TT més possible és l’error.

No sembla que el sexe condicioni l’encert, tot i que no podríem demostrar-ho ja que la mostra no ha estat equitativa en aquest camp. En canvi, l’edat sí que sembla condicionar l’encert, ja que dels 9 errors que s’ha produït 6 són del grup més jove (el de 13 anys). Tot i així no podem afirmar que l’edat condiciona a l’encert perquè no hem fet una distinció de grups d’edat i no tenim una mostra prou gran per comparar i perquè poden existir altres factors que no hem tingut en compte (com pot ser els coneixements previs) que provoquin confusió.

D’altra banda, vam fer una anàlisi qualitativa del contingut de les converses realitzades pels subjectes i dels motius d’encert. Les principals raons dels subjectes que encerten són que el xatbot és evasiu o no entén directament. Considerem que el xatbot, en força ocasions, ha tingut un comportament molt similar a l’humà. L'expectativa inicial entorn al comportament del nostre xatbot ha sigut superada completament.

**Limitacions del nostre estudi:**

En la realització dels tests vam haver de modificar el disseny d’aquest degut a que vam observar que el xatbot era incapaç d’escriure per la via de comunicació utilitzada (Discord). Això ens va fer prendre la decisió d’incloure una tercera persona, un cop iniciat l’experiment, encarregada de transcriure els missatges del xatbot (tot i que en un inici s’esperava que el xatbot contestes per ell sol).

En el procés d’entrenament del xatbot vam constatar una gran dificultat a l’hora de pensar quines podrien ser les respostes i les intencions necessàries perquè el xatbot fos capaç de simular un comportament humà. Vam objectivar la nostra mancança de recursos (tant en relació al temps que podíem invertir en el procés d’entrenament com en altres qüestions tècniques i inclús lèxiques) per ser capaços de pensar més possibilitats de converses que permetessin al xatbot tenir un ventall de respostes pràcticament infinites. També creiem que el fet de limitar a la persona a utilitzar només les respostes del xatbot ha afavorit positivament el grau d'error dels tests, ja que tenim el dubte que sense aquesta limitació, el xatbot hagués estat capaç de superar el TT.

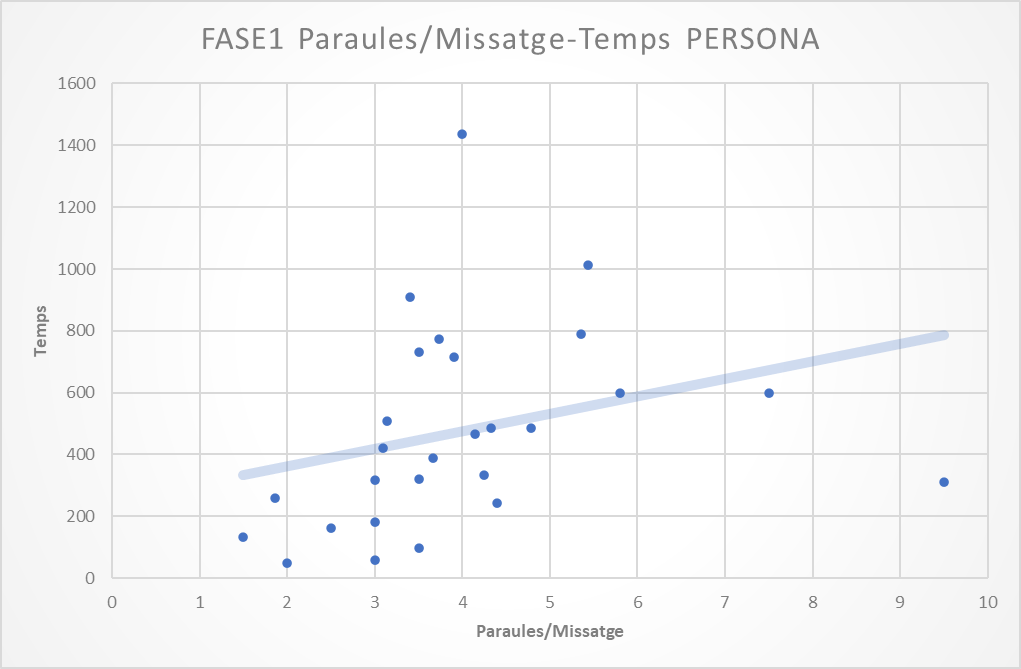
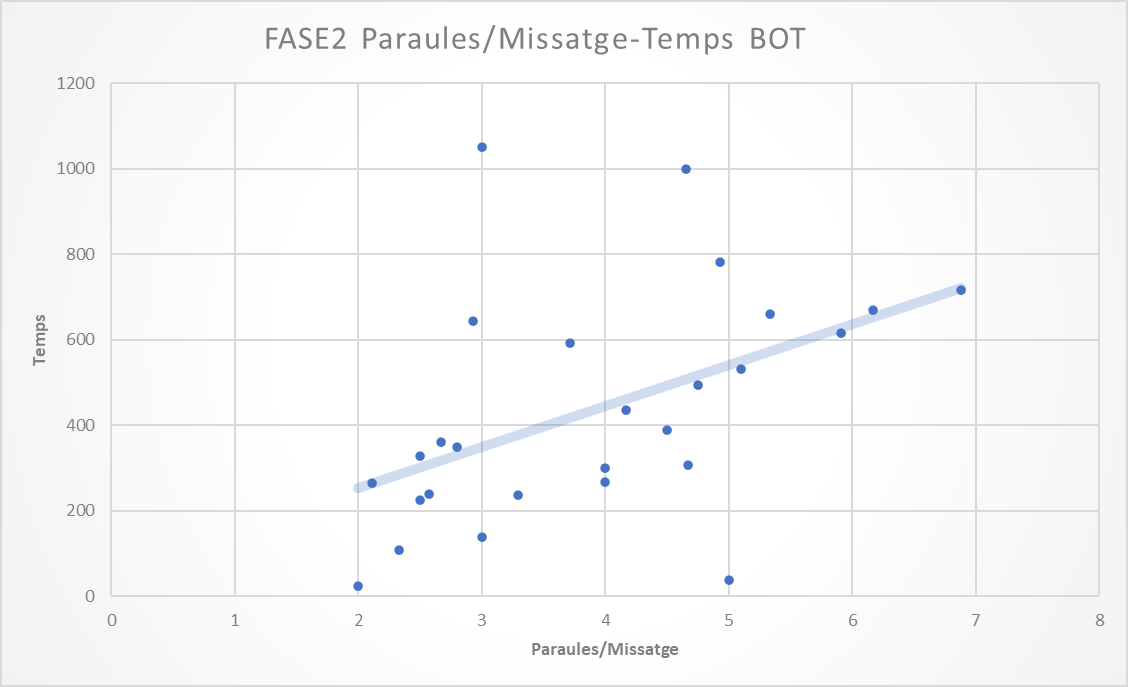
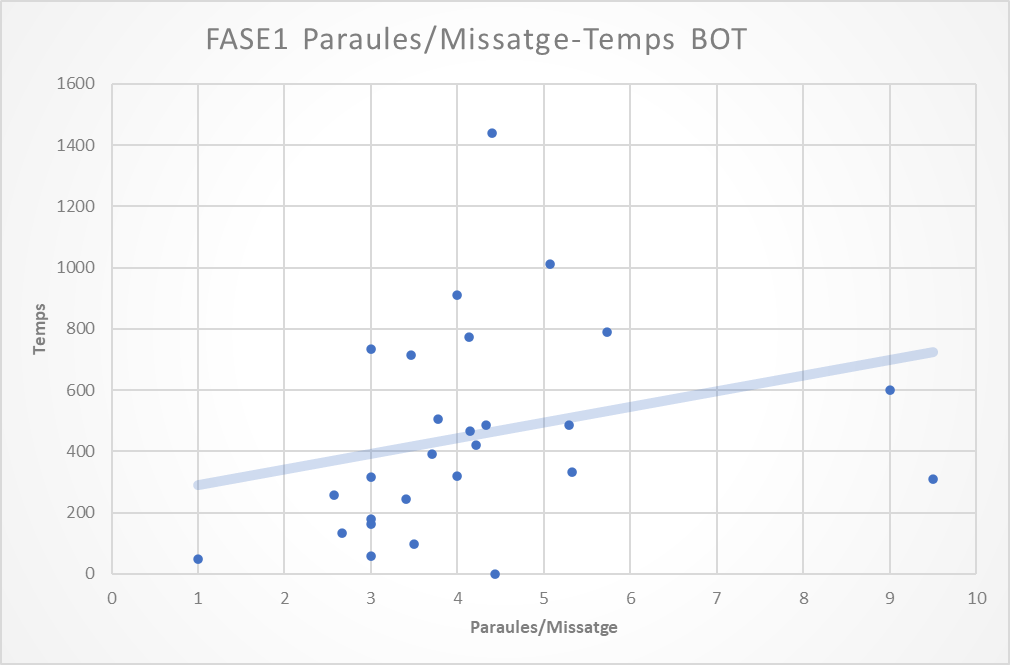
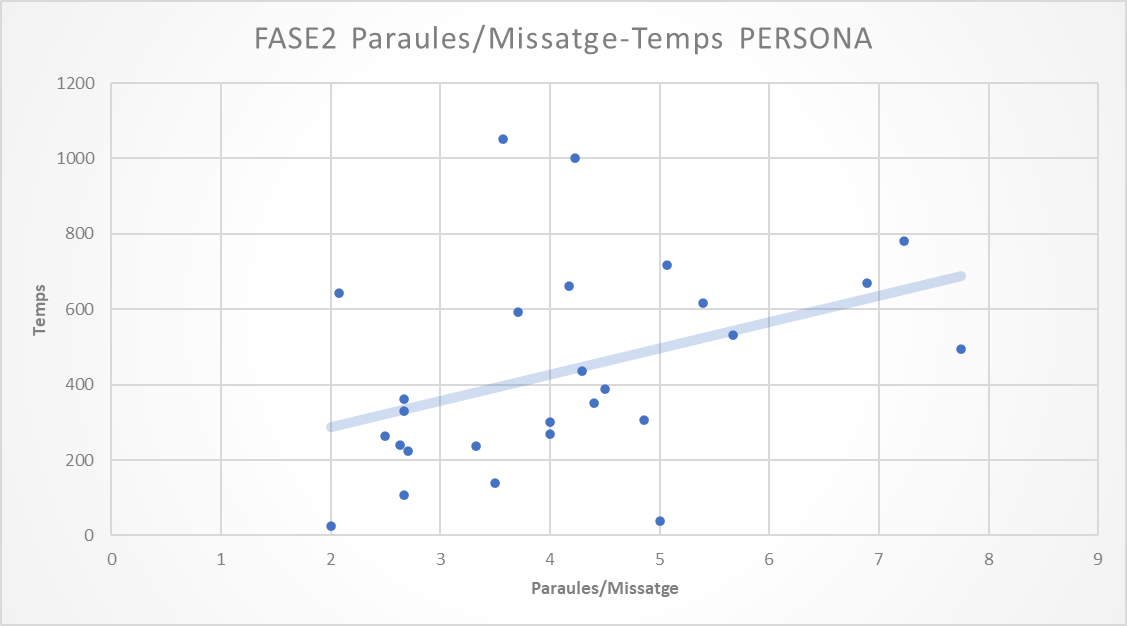
També pensem que podríem haver afinat alguns detalls en l’entrenament del xatbot. Per exemple: En casos on els subjectes saludaven amb un “*buenas”* el xatbot responia amb un “*gracias”*… A més es podria haver acotat el test amb la prohibició dels accents i de caràcters símbols (^`:\*,;) ja que el xatbot ha mostrat complicacions amb la seva utilització.

Considerem també que hem fet servir un nombre màxim de missatges massa gran. Intuïm que amb 20 missatges per conversa i amb la realització de dues proves consecutives, possiblement els tests siguin massa tediosos i demanin més temps de realització. Si el nombre màxim de missatges es reduís, per exemple, a la meitat creiem que la realització dels tests seria més àgil i els subjectes haurien d’utilitzar una estratègia més eficient per aconseguir resoldre’ls.  
.

**CONCLUSIÓ:**

* Perspectiva de futur de la recerca amb IA
* Es pot entrenar en benefici d’un objectiu comú/quin sentit té aquesta experimentació
* quin ús futur tindria el que hem fet? té sentit el que hem descobert amb el nostre experiment? (és una tecnologia que pot fer molt de mal o molt de bé segons com s’utilitzi)
* Les sensacions subjectives en relació a tenir un treball de recerca amb una part experimental
* Les preguntes que se’ns han generat mentre fèiem la recerca

**ANNEXS:**



La relació entre paraules per missatge enviades al xatbot i a la persona i el temps a ambdues fases. En aquest cas la relació era casi nul·la, amb un coeficient de Pearson de 0.27, 0.29(FASE1), 0.48 i 0.39(FASE2) respectivament.

Consentiment Informat

Full d’Informació:

Test de Turing (Lucas Díaz Lopez, Ignacio Antón Martín, Institució Cultural del C.I.C):

Si us plau, llegeix acuradament aquest document de consentiment abans de decidir-te a participar en aquest estudi.

Objectiu:

L’objectiu d’aquesta investigació és realitzar un *Imitation game* que té per objectiu mesurar la capacitat d’una màquina per fer-se passar per una persona humana.

Implicacions de la participació:

Et demanarem que participis en dues proves de forma consecutiva, és a dir una darrere de l’altra. Les dues proves seguiran les mateixes instruccions i objectius.

Per a realitzar les proves hauràs d’utilitzar un dispositiu electrònic (ordinador, mòbil, tauleta...) Un dels investigadors principals de l’estudi serà present al teu costat durant tot el temps que duri la realització de les proves.

Les proves consisteixen a mantenir dues converses simultànies de forma telemàtica, Les converses estaran identificades amb la lletra A o B. L’idioma utilitzat serà el castellà.

Una de les converses la faràs amb una persona humana i l’altra la faràs amb una màquina. Tu no sabràs a qui correspon cada conversa. Totes dues converses es produiran simultàniament i les veuràs reflectides a l’hora al teu dispositiu.

El teu objectiu serà identificar quina conversa s’està produint amb la màquina i quina amb la persona humana. Per fer-ho pots utilitzar totes les estratègies comunicatives que creguis necessàries (per exemple: fer preguntes, fer servir paranys o enganys que et permetin esbrinar qui és l’humà i quina la màquina).

Només tens una única limitació que són el nombre de missatges, ja que només podràs escriure un màxim de 20 missatges. En cas que esbrinis qui és qui abans dels 20 missatges, pots resoldre el joc en el moment que vulguis sense necessitat d’utilitzar tots els missatges. Quan identifiquis quina és la conversa realitzada per la persona humana i quina per la màquina hauràs de comunicar-li-ho a l’investigador que t’acompanyi relacionant A i B amb persona humana o màquina.

Durant el desenvolupament de l’experiment, l’investigador principal calcularà el temps invertit en la resolució del joc.

El resultat del teu joc te’l comunicarem a través d’un missatge de correu electrònic un cop haguem analitzat les dades de tots els participants.

Durada:

La durada màxima del joc serà aproximadament de 15 minuts.

Riscos i beneficis:

No es contemplen riscos en la teva participació.

Compensació:

En aquest cas no està prevista cap compensació per participar-hi.

Confidencialitat:

Si decideixes participar-hi, la teva identitat es mantindrà confidencial i només els membres de l’equip de recerca tindran accés a les dades del projecte. S’utilitzarà un codi alfanumèric per identificar a cada participant que només coneixeran els investigadors principals.

Voluntarietat:

La participació en aquest estudi és completament voluntària.

Dret a retirar-te de l’estudi:

Tens el dret a retirar-te de l’estudi en qualsevol moment sense donar explicacions només informant als investigadors de l’estudi. No hi haurà cap conseqüència negativa de la teva retirada voluntària de l’estudi.

Persona de contacte:

En cas de dubte o consulta pots contactar amb:

Lucas Díaz López [correo.lucas.ldl@gmail.com](mailto:correo.lucas.ldl@gmail.com)

Ignacio Antón Martín [ignacioanma05@gmail.com](mailto:ignacioanma05@gmail.com)

Full de Consentiment:

Test de Turing (Lucas Díaz Lopez, Ignacio Antón Martín, Institució Cultural del C.IC)

Consentiment:

· He llegit la informació sobre el projecte de recerca i he tingut l’oportunitat de fer preguntes, les quals se m’han respost satisfactòriament.\_\_

· Estic d’acord a participar-hi i he rebut una còpia d’aquest consentiment.\_\_

Nom i cognoms del participant \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signatura \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Data: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Investigador/a que m’ha informat:

Signatura \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Data:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Cronograma:**

**BIBLIOGRAFIA:**

1.

TURING AM. COMPUTING MACHINERY AND INTELLIGENCE (TURING’S ARTICLE). Mind. 1 octubre 1950;LIX(236):433-60.

3.

Domain [Internet]. [citat 19 setembre 2022]. Disponible a: <https://rasa.com/docs/rasa/domain/>

4.

Explained: Neural networks [Internet]. MIT News | Massachusetts Institute of Technology. [citat 19 juny 2022]. Disponible a: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>

5.

Fallback and Human Handoff [Internet]. [citat 19 setembre 2022]. Disponible a: <https://rasa.com/docs/rasa/fallback-handoff/>

6.

Chowdhary KR. Fundamentals of Artificial Intelligence [Internet]. New Delhi: Springer India; 2020 [citat 16 juny 2022]. Disponible a: <http://link.springer.com/10.1007/978-81-322-3972-7>

7.

Asghar S. GPT3 [Internet]. Demo e. [citat 1 maig 2022]. Disponible a: <https://gpt3.website>

8.

Intents & Entities: Understanding the Rasa NLU Pipeline [Internet]. Rasa. [citat 19 setembre 2022]. Disponible a: <https://rasa.com/blog/intents-entities-understanding-the-rasa-nlu-pipeline/>

9.

Is the Turing Test Obsolete? - ExtremeTech [Internet]. [citat 17 juny 2022]. Disponible a: <https://www.extremetech.com/computing/318767-is-the-turing-test-obsolete>

10.

Boden MA. Mind as machine: a history of cognitive science. Oxford: Clarendon Press; 2006. 2 p.

11.

NLU Training Data [Internet]. [citat 19 setembre 2022]. Disponible a: <https://rasa.com/docs/rasa/nlu-training-data/>

12.

Rasa Playground [Internet]. [citat 19 setembre 2022]. Disponible a: <https://rasa.com/docs/rasa/playground/>

13.

Rules [Internet]. [citat 19 setembre 2022]. Disponible a: <https://rasa.com/docs/rasa/rules/>

14.

Stories [Internet]. [citat 19 setembre 2022]. Disponible a: <https://rasa.com/docs/rasa/stories/>

15.

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. :19.

16.

Test de Turing: ¿pueden las computadoras sustituir a los humanos? [Internet]. [citat 20 juny 2022]. Disponible a: <https://www.becas-santander.com/es/blog/test-de-turing.html>

17.

The Google engineer who thinks the company’s AI has come to life. Washington Post [Internet]. [citat 17 juny 2022]; Disponible a: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>

18.

What is AI? Here’s everything you need to know about artificial intelligence [Internet]. ZDNet. [citat 19 juny 2022]. Disponible a: <https://www.zdnet.com/article/what-is-ai-heres-everything-you-need-to-know-about-artificial-intelligence/>

1. Test de Turing: ¿pueden las computadoras sustituir a los humanos? [citat 20 juny 2022]. Disponible a:<https://www.becas-santander.com/es/blog/test-de-turing.html> [↑](#footnote-ref-2)
2. Is the Turing Test Obsolete? - ExtremeTech [Internet]. [citat 17 juny 2022]. Disponible a:<https://www.extremetech.com/computing/318767-is-the-turing-test-obsolete> [↑](#footnote-ref-3)
3. Chatbot Trends Report 2021 [citat 20 juny 2022]. Disponible a: <https://chatbotsjournal.com/chatbot-trends-report-2021-b15479c404e4> [↑](#footnote-ref-4)
4. Asghar S. GPT3 [Internet]. Demo e. [citat 1 maig 2022]. Disponible a: [https://gpt3.website](https://gpt3.website/) [↑](#footnote-ref-5)
5. The Google engineer who thinks the company’s AI has come to life. Washington Post [Internet]. [citat 17 juny 2022]; Disponible a:<https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/> [↑](#footnote-ref-6)
6. *Francois Chollet*: Investigador d’IA a Google i creador de Keras (bibilioteca de software machine-learning) [↑](#footnote-ref-7)
7. What is AI? Here’s everything you need to know about artificial intelligence [Internet]. ZDNet. [citat 19 juny 2022]. Disponible a: <https://www.zdnet.com/article/what-is-aiheres-everything-you-need-to-know-about-artificial-intelligence/> [↑](#footnote-ref-8)
8. Explained: Neural networks [Internet]. MIT News | Massachusetts Institute of Technology. [citat 19 juny 2022]. Disponible a:

   <https://news.mit.edu/2017/explainedneural-networks-deep-learning-0414> [↑](#footnote-ref-9)
9. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: LargeScale Machine Learning on Heterogeneous Distributed Systems. [citat 18 juny 2022] Disponible a: <http://download.tensorflow.org/paper/whitepaper2015.pdf> [↑](#footnote-ref-10)
10. Introduction to Rasa Open Source. [citat 18 juny 2022] Disponible a:

    <https://rasa.com/docs/rasa/> [↑](#footnote-ref-11)
11. NLU Training Data. [citat 18 juny 2022] Disponible a:

    <https://rasa.com/docs/rasa/nlu-training-data/> [↑](#footnote-ref-12)
12. Intents & Entities: Understanding the Rasa NLU Pipeline. [citat 18 juny 2022] Disponible a:

    [Intents & Entities: Understanding the Rasa NLU Pipeline | The Rasa Blog | Rasa](https://rasa.com/blog/intents-entities-understanding-the-rasa-nlu-pipeline/) [↑](#footnote-ref-13)
13. Rasa Playground. [citat 18 juny 2022] Disponible a:

    <https://rasa.com/docs/rasa/playground>

    Training Data. [citat 18 juny 2022] Disponible a:

    <https://rasa.com/docs/rasa/training-data-format> [↑](#footnote-ref-14)
14. YAML és un llenguatge de serialització de dades que sol utilitzar-se en el disseny d'arxius de configuració. [↑](#footnote-ref-15)
15. Domain. [citat 13 juliol 2022] Disponible a:

    <https://rasa.com/docs/rasa/domain/> [↑](#footnote-ref-16)
16. NLU Training Data. [citat 18 juny 2022] Disponible a:

    <https://rasa.com/docs/rasa/nlu-training-data/> [↑](#footnote-ref-17)
17. Rules. [citat 13 juliol 2022] Disponible a:

    <https://rasa.com/docs/rasa/rules>/ [↑](#footnote-ref-18)
18. Stories. [citat 13 juliol 2022] Disponible a:

    <https://rasa.com/docs/rasa/stories/> [↑](#footnote-ref-19)
19. FallBack. [citat 15 juliol 2022] Disponible a:  
    <https://rasa.com/docs/rasa/fallback-handoff/> [↑](#footnote-ref-20)
20. TURING AM. COMPUTING MACHINERY AND INTELLIGENCE (TURING’S ARTICLE). Mind. 1 octubre 1950;LIX(236):433-60. [citat 13 juny 2022] Dispoible a: <https://academic.oup.com/mind/article/LIX/236/433/986238> [↑](#footnote-ref-21)
21. TURING AM. COMPUTING MACHINERY AND INTELLIGENCE (TURING’S ARTICLE). Mind. 1 octubre 1950;LIX(236):433-60. [citat 13 juny 2022] Dispoible a: <https://academic.oup.com/mind/article/LIX/236/433/986238> [↑](#footnote-ref-22)