

Declaració de treball original (no plagiat) de l'estudiant

Jo, Xavi Rubio, declaro que per realitzar aquest lliurament m'he inspirat, he mirat, i m'he informat però no he copiat ni plagiat ningun contingut ni un document d'un company ja sigui actual o passat.

Introducció a la ciència de dades

PAC3: Detectar el dilema ètic i els riscos de no governar les dades

Pregunta 1 (30% puntuació)

1.1.

La primera seria la transparència en els procediments on qualsevol entitat o individu estigui involucrat, i, la segona el coneixement necessari precís i actualitzat de la legislació que afecta actualment al *big data* on també s'hauria de saber l'origen i la resolució de qualsevol conflicte generat. Els dos procediments anteriors recolzats d'una estimació oberta de les implicacions a llarg termini podrien facilitar una alineació estratègica.

Ex. Tenim una empresa que actualment no aplica aquests principis ètics, i un dia tenen qualsevol problema legal mediàtic on s'ha fet eco a les xarxes socials i per conseqüència s'han generat 'teories' falses i on ho tenen que resoldre per la via judicial, si haguessin aplicat el segon procediment tot això que no tenien i per lo qual no sabien se li complicarà compromentent a la empresa a grans despeses econòmiques.

1.2.

Al cap i a la fi el primer valor ètic que hauria de assumir una organització te directament a veure amb els seus clients, ja que, que es una empresa sense clients... I aquests clients son persones que es basen en la confiança, i que millor que fer la confiança empresa-client o B2C lo mes proper possible, es a dir, donar una total transparència de la utilització i/o emmagatzematge de les dades. Com es pot veure^[1] empreses com Microsoft i Samsung tenen un aprovat molt just en les categories de B2B i B2C, això el que fa es a nivell de consumidor millorar la relació amb aquesta empresa ja que saps que no utilitzaran les teves dades en la teva contra, sinó que les utilitzaran com a millora de l'experiència com usuari d'aquestes empreses. També si la organització te un excel·lent coneixement de la legislació que afecta a les dades podrien reduir el cost de qualsevol problema que pugui sorgir, a mes a mes, el fet de saber l'origen del qual ve aquest problema podria evitar inclús una resolució per la via judicial i nomes quedar-se en 'un problema resolt'.

Bibliografia

- [1] M. Tilves, «Poniendo nota a las tecnológicas en innovación, transparencia y algo más [Infografía],» 28 octubre 2015. [En línia]. Available: <https://www.silicon.es/poniendo-nota-a-las-tecnologicas-en-innovacion-transparencia-y-algo-mas-infografia-90143>.

1.3.

Els punts de decisió ètica son els següents: recerca, que tracta sobre el punt on la empresa ha de saber qui es, el que vol fer i perquè, ja que així el que es fa es donar una identitat a la empresa el qual ajudarà a la recerca de la prena de decisions ètiques; el anàlisis, que sabent i tenint la identitat de la empresa el que es faria seria identificar una decisió tenint en compte la identitat i organització que mes s'adeqüés al problema en concret on s'hauria de testear la seva validesa; articulació, el punt on s'haurien d'aplicar els principis ètics de la prena de decisió que la organització hagi escollit com a preferida; i acció, que seria el resultat de aplicat tots els punts anteriors a la resolució final de la organització. I els termes mes important en el moment d'atacar un punt de decisió ètica son: la intenció, la qual es refereix a les intencions finals de qualsevol mitja ja sigui directe o indirecte de les dades, com per exemple: la filtració d'un virus informàtic al servidor de BBDD on estan emmagatzemades les dades i la intenció d'aquest es publicar-les; la seguretat, es el que la pròpia organització utilitza per establir les mesures de seguretats establertes per llei o l'exigència interna de la pròpia organització, com per exemple, utilitzar sistemes operatius actualitzats, restringir l'accés a les instal·lacions...; la probabilitat, es el resultat de un anàlisis de seguretat per veure segons la seguretat actual de la organització quina probabilitat n'hi ha de que tinguin problemes de seguretat; la agregació, també la pròpia organització s'ha de fer responsable de la probabilitat de problemes derivats a les dades que actualment emmagatzemen; la responsabilitat, es refereix a la responsabilització de la empresa com a conseqüència de una presa de decisions feta prèviament, com per exemple, prendre una decisió la qual ha perjudicat a la empresa; la identitat, quines dades utilitza la organització per a cada individu les quals s'utilitzen per identificar al individu com per exemple, les direcció IP o el tipus de dispositiu que si s'ajunten es pot fer una cerca concreta d'un dispositiu; la propietat, indica sobre l'estat dels drets d'ús que aplica en cada punt de la cadena, per exemple, que un usuari pugui saber fins a quin punt se li pot proporcionar les seves dades; el benefici, qualsevol contribució positiva de l'organització cap a la l'usuari o de la pròpia organització com a millora de la mateixa, com per exemple, l'ús de les dades com a finalitat de millorar el rendiment de la empresa i dels seus treballadors; i el dany, on podria ser dany cap a la pròpia organització, sobre la identitat, la privacitat, la intimitat o la reputació les quals deriven d'un accés a dades especifiques, com per exemple, si una organització representa un icona de seguretat i principalment nomes te problemes de seguretat danya a la identitat i la reputació de l'organització.

Pregunta 2 (35% puntuació)

Per primer poder analitzar aquesta eina ^[2], hem de veure com funciona una mica mes en concret, sabem que agafa la informació de repositoris de GitHub, però si veiem en la pagina oficial de Copilot veiem que el que diuen es que aquesta eina esta entrenada a partir de repositoris públics, això es un punt de partit important y amb el que treballarem. Si partim de la premissa de que utilitza repositoris totalment públics podem dir que utilitza codis open-source, que son codis oberts els quals qualsevol persona por accedir-hi sense cap problema i fer amb ells el que vulguin, ja sigui editar-los i fins i tot comercialitzar-los. Per acabar d'especificar una mica mes les restriccions dels repositoris de GitHub hauríem de tenir en compte que cada repositori te una llicencia ^[3] la qual el propi autor pot dir si vol que el seu repositori tingui alguna llicencia o no hi apliqui cap o si hi vol poder escollir entre algunes opcions, entre aquestes llicencies n'hi ha d'algunes que el que fan es que encara que sigui un repositori públic no es pugui ni comercialitzar ni utilitzar a cap lloc, en aquest cas l'organització d'aquesta eina hauria de fer que amb l'aparició d'aquesta en crear un repositori pogués simplement marcar una casella com a que vol que el seu codi s'utilitzi per a dades d'entrenament, o notificar que en tots els repositoris públics o els que tenen una llicencia en concret s'utilitzaran com a dades d'entrenament.

Tenint en compte tot això poder dir que d'acord amb el conseqüencialisme aquesta eina esta creada per fer el be, per a ajudar a la gent ja que aquesta organització te una quantitat de dades molt útils les quals com ja hem dit son publiques, però aquí pareixen les teories no teleològiques, com a representant únic, la deontologia, la qual utilitzen ja que com hem vist encara que podent utilitzar la totalitat dels seus repositoris nomes utilitzen els quals son públics, es a dir, que el reu autor ha decidit que el seu codi sigui en la seva forma open-source. Ara que tenim ja analitzat les teories ètiques a les quals esta subjecta aquesta eina podem avaluar com compleix aquest objectiu que hem avaluat prèviament. Aquí s'aplica el principi de beneficència el qual si partim de les dues regles bàsiques (1. No causar cap dany; 2. Maximitzar els possibles beneficis, disminuint els possibles danys) podem veure que a grans trets es compleix, ja que si l'autor posa te el repositori públic o no, però si per altra banda com hem comentat entrem amb el tema de les llicencies que hi podem escollir com a usuaris, la cosa canvia. Ja que un entrenament o per el contrari un tros de codi el qual ha extret d'un repositori amb llicencia i que una empresa utilitza per a la seva aplicació ja no es compliria el principi de beneficència en la seva totalitat. També compleix el principi d'autonomia, ja que l'usuari en tot moment podria modificar o escollir l'estat del seu repositori, i, en canvi en el principi de no maleficència podem veure com encara que les llicencies estiguin posades aquesta eina segueix agafant aquests codis per a entrenar i aquí entra el principi de justícia on podem veure que el dilema moral d'utilitzar un codi amb llicencia que utilitza es el denominat 'principi de diferencia', el qual estableix que en aquest cas l'únic que fa es entrenar per a un problema comú, i el qual ajudaria a la totalitat de les persones que utilitzen aquesta eina i no nomes a utilitzar-la en benefici parcial o propi.

Ara que ja tenim justificat els principis morals als quals afecta aquesta eina, un altre problema el qual tindria es la propietat de les dades que genera aquesta eina. Primer hem de diferenciar entre propietat i pertinença, així podríem fer una definició molt mes exacta. Tenint en compte que el que fa es generar codi que com veiem en els links de l'enunciat ho fa de una forma quasi exacta al codi que treu dels repositoris podríem dir que aquest codi com porta una referencia de l'usuari del que l'ha fer servir, encara que no sigui 100% exacte al codi de l'usuari aquest codi

generat te una propietat que es la del usuari que utilitza l'eina ja que es generada automàticament, però sigui exacte o no la pertinència segueix sent de l'autor d'aquell codi, ja que si algú fa qualsevol cosa que tu utilitzes segueix sent de l'autor principal encara que tu ara l'utilitzis sigui per a qualsevol cosa sempre i quant ho deixis tal qual esta, ja que si el modifiques significativament com es de la teva propietat podries utilitzar-lo en el teu benefici. Aquesta eina pot tenir errors ja que no es perfecte, perquè com surt a l'enunciat encara es una primera versió la qual pot tenir errors i no precisament petits com indica a PWV Consultants^[4]. En aquest punt i amb problemes tan greus ja que no només són problemes de codi els quals tenen errors sintàctics sinó que escriu coses les quals discriminen a grups vulnerables si que s'hauria d'actuar d'alguna manera o fer algun canvi intern per a així poder fer un eina totalment parcial i que tingui només l'objectiu principal sense fer cap mal a cap grup social. El propietari de qualsevol repositori no hauria de estar obligat de cap mena en la utilització d'aquesta eina i sempre hauria de ser opcional per a qualsevol usuari que la vulgui utilitzar, així que cada persona l'utilitza per al seu propi benefici el qual esta suportat per la organització darrera d'aquesta eina on hem vist que manté uns principis morals correctes així que mentre tot segueix així i continuïn millorant-la no seria necessari cap mena de supervisió per cap mena d'entitat o organització fora de els creadors i els usuaris de l'eina.

Bibliografia

- [2] «Your AI pair programmer,» GitHub, 2021. [En línia]. Available: <https://copilot.github.com/>.
- [3] «Licensing a repository,» GitHub, [En línia]. Available: <https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/licensing-a-repository#disclaimer>.
- [4] P. Vanlperen, «Questions Around Bias, Legalities in GitHub's Copilot,» PWV Consultants, 14 July 2021. [En línia]. Available: https://www-pwvconsultants-com.translate.goog/blog/questions-around-bias-legalities-in-githubs-copilot/?_x_tr_sl=auto&_x_tr_tl=es&_x_tr_hl=en-US&_x_tr_pto=op.

Pregunta 3 (35% puntuació)

3.1.

Per primer saber com aplicar un cicle de vida de les dades als cotxes autònoms primer hauríem de descriure el valor de les dades les quals com ajudarien al negoci, per exemple si tenim un cotxe autònom podem registrar absolutament tota la informació que esta el voltant d'aquest, on aquesta informació ajudaria a la millora del programa del cotxe i la qual es podria aplicar a altres millores o projectes dins de l'organització per a poder millorar la seguretat de conducció. Encara que es pot registrar dades de tot el que esta al voltant del cotxe això suposaria una gran

inversió que pot ser no son capaços de fer així que s'hauria de escollir quins dispositius utilitzar i definir la utilització d'aquests. Una vegada establertes aquestes fases ara hauríem de organitzar un mar apropiat, en aquest cas l'organització utilitzaria totes les dades generades per l'usuari, i l'usuari podria tenir accés a les dades que ell mateix genera, però no a les generades per l'organització gracies a l'usuari. Una vegada concretat aquesta fase i amb tots els permisos d'accés a les dades de l'organització ja podríem començar a recollir les dades i organitzar un cicle de vida. Com ja he dit anteriorment depenent dels dispositius escollits capturarem les dades de diferent manera, però sempre s'hauria de seguir el mateix procediment, processar les dades entendre-les, fer-ne us de les dades i per acabar donar-li un us quan la dada ja esta 'morta' per a l'organització.

3.2.

Primer hem de desenvolupar la definició i la utilitat que li donarem a les dades que generarem dins de la pròpia organització, ja sigui per a que volem aquestes dades, com les farem servir, quines utilitats i impacte tindran dins de l'organització el fet de tenir-les... Una vegada s'entén el que faran les dades a l'organització hem de fer un full de ruta el qual impliqui definir per a que es faran servir les dades, com per exemple podria ser alguna cosa relacionada amb la millora dels components de recollida de dades o un projecte diferent però que utilitzi les dades recollides com per exemple la creació d'un model de cotxe específic o una millora cap a l'usuari. Això implica treball el qual s'ha d'organitzar així que es planifica concretant els diferents projectes estudiats i fent un càlcul de pressupost tenint el compte els costos humans, i financers que pot compondre realitzar el projecte, on també s'hauria es defineix el component de l'Arquitectura de dades. Una vegada feta tota la part de definició ara toca començar a fer el disseny concret el qual s'utilitzarà per a cada projecte on els components son l'emmagatzematge de dades, dades mestres i referencia, el disseny i models de dades, la integració i interoperabilitat i la intel·ligència de negoci i magatzems de dades. Una vegada dissenyat i testejat el model que s'utilitzarà toca implementar-ho al mon real on començarà a fer el que ha estat descrit anteriorment, on utilitzarà el contingut i document. Una vegada el model implementat es passa a la fase de governar les dades on implementes l'organització de les dades i identifiqués l'accessibilitat d'aquestes dins i fora de l'organització amb el component Seguretat de dades. I per acabar una vegada ja reculls les dades les dones un us a futur.

3.3.

Una organització o empresa si vol aconseguir reptes i superar-se com a empresa i als seus competidors necessita complementar una bona organització i gestió de les dades complementat amb una bona governança de dades, ja que mitjançant com gestionis les dades tindràs un rendiment diferent i aquí es on entra el *data governance*, sempre hauràs de tenir complementació entre aquests dos camps, ja que si vols poder millorar els dispositius que utilitzes en aquest cas per als cotxes autònoms, els que no faràs es gestionar la velocitat en la qual el cotxe accelera de 0-100 km/h, perquè mai arribaries a poder millorar eficient. En el cas dels cotxes autònoms la gestió d'aquests dos camps es imprescindible, i no nomes utilitzar-los, sinó que fer-ho lo millor possible utilitzant els recursos necessaris o els que es tinguin disponibles. Continuant amb aquest cas l'aplicació mes idònia que es podria fer ja que sobretot el que s'hauria de buscar es la seguretat tant del conductor com de la gent dintre del cotxe i la

de fora d'aquest seria tota la millora dels dispositius que recullen les dades del seu voltant, la utilització del conductor del vehicle, sabent la reacció del conductor cap a qualsevol problema, a mes a mes de la millora del software amb el qual el cotxe reaccionarà als imprevistos de l'exterior.