

Guia per executar projectes d'Analítica Avançada i IA

Octubre 2024



Next Generation
Catalunya



**Generalitat
de Catalunya**

0

Control de canvis del document

0. Control de canvis del document

Versió	Data	Autor	Comentaris
v1.01	01/10/2024	Equip STA	Primera proposta de la guia presentada a la DGSEC y CTTI
v1.02	08/10/2024	Equip STA	Modificacions sorgides de la sessió de revisió conjunta amb l'equip de la DGSEC i CTTI.
v1.03	17/10/2024	Equip STA	Modificacions sorgides arran de comentaris en la presentació per part dels equips de DGSEC i CTTI.

1

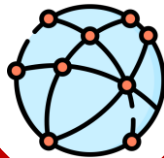
Introducció

Objectiu

Objectiu general

Fomentar la transformació digital i la presa de decisions basades en dades dels Departaments de la Generalitat de Catalunya a través d'agilitzar i garantir la qualitat del desenvolupament de casos d'ús d'analítica avançada i IA

Objectius específics



Establir una guia pràctica
que garanteixi una visió
holística al llarg del cicle de
vida dels projectes



Garantir que els
responsables dels sistemes
d'analítica avançada i IA
comprenguin les seves
responsabilitats i les fases
del projecte

2

Organització: Rols

2. Organització: Rols

Rol	Àmbit	Descripció
<u>Responsable d'impacte algorímic</u>	Central	— Identifica riscos algorísmics i proposa mesures mitigadores. — Aplica criteris ètics establerts pel Comitè d'Ètica de les Dades i la Política d'ètica de les dades. — Escala riscos nous i compta amb suport extern expert en avaluació algorítmica.
<u>Responsable del Registre</u>	Central	— Encarregat del govern centralitzat del Registre (DGSEC amb la col·laboració del CTTI). — Centralitza la informació necessària per disposar de la visió global dels sistemes d'IA i algorismes a l'Administració. — Assegura el funcionament i actualització del Registre. — Promou canvi cultural mitjançant activitats de difusió i formació per a tots els perfils usuaris. — Valida altes i canvis, i realitza revisions periòdiques de la informació del Registre.
<u>Responsables de protecció de dades</u>	Federat	— Vetllen per la protecció de dades al llarg de tot el cicle de vida del sistema d'IA/algorisme des de la seva recopilació fins a la seva eliminació. — Supervisen la recol·lecció i processament de les dades personals. — Gestionen el consentiment i drets dels usuaris. — Acompanyen els responsables de sistemes d'IA en la implementació i gestió de riscos.
<u>Responsable de ciberseguretat</u>	Federat	— Vetllen per la protecció de la infraestructura i actius digitals enfront d'amenaques i vulnerabilitats. — Asseguren la confidencialitat, integritat i disponibilitat dels sistemes i les dades. — Implementen polítiques de seguretat. Monitoren amenaces i prevenen atacs. — Acompanyen els responsables de sistemes d'IA en la implementació i gestió de riscos.

2. Organització: Rols

Rol	Àmbit	Descripció
<u>Referent Govern de les dades i IA</u>	Federat	– Coneix els processos, dades i els principis de govern. – Supervisa que es compleixin les directrius i estàndards de govern en tot el departament, tant estratègicament com operativament. – Coordina amb responsables funcionals, i és el nexa entre els òrgans centrals i departamentals. – Gestió de Riscos: Assessora i dona suport als responsables del sistema d'IA en la gestió de riscos. – Registre: segura que els responsables d'IA usin el Registre i que la publicació de dades obertes es faci d'acord amb els principis establerts.
<u>Referent Tècnic Govern de les dades i IA</u>	Federat	– Perfil tecnològic que executa i coordina accions per al govern de les dades. – Es coordina amb l'arquitecte de dades i els responsables tècnics dels sistemes d'informació. – Transmet directrius del govern central a les àrees TIC i viceversa. – Col·labora en la identificació de noves oportunitats i casos d'ús.
<u>Responsables de la unitat</u>	Federat	– Perfil funcional que coneix el seu àmbit de negoci. Alinea el projecte amb els objectius de negoci i les necessitats dels usuaris. – Treballa de forma coordinada amb el referent de govern de les dades per aplicar els principis de govern de les dades. – Responsable directe d'aplicar i aterrar el Marc de gestió de riscos al seu sistema d'IA. – Omple la informació dels camps del Registre i la documentació annexa. – Pren el control del projecte un cop finalitzat, assegurant la seva continuïtat i evolució.
<u>Cap del projecte d'IA/AA</u>	Federat	– Perfil tècnic, encarregat de gestionar i coordinar l'equip tècnic (arquitecte, enginyer de dades, científics de dades, etc.). – Supervisa el desenvolupament tècnic i la qualitat de la solució. – S'enfoca en els aspectes tècnics del projecte, com la infraestructura, models d'IA, i la resolució de problemes tècnics. – Una vegada finalitzat el projecte, aquest rol desapareix.

2. Organització: Rols

Rol	Àmbit	Descripció
<u>Arquitecte de dades</u>	Federat	– Coneix les especificacions tècniques per definir correctament les dades en la infraestructura tecnològica, de manera que s'aconsegueixi un mapa complet que representi com i on estan localitzats, i també com s'interrelacionen. – Alinea l'estratègia tecnològica amb la funcional, aportant coneixement tècnic sobre les diferents solucions per a la gestió de dades. – Dissenya l'arquitectura per al desplegament de sistemes d'IA i aporta coneixements tècnics sobre les solucions.
<u>Administrador de sistemes</u>	Federat	– Administra i controla els sistemes que són font d'informació. – Implementa controls de qualitat, seguretat i privacitat en l'àmbit tecnològic, alineant les necessitats funcionals amb els sistemes tècnics.
<u>Enginyer de dades</u>	Federat	– Implementa processos per a l'adquisició i migració de dades. – Identifica les dades necessàries per a cada cas d'ús. – Desenvolupa processos de tractament de dades. – Gestiona l'accés i provisió de dades per als Data Scientists. – Monitoritza els processos d'extracció i càrrega de dades, resolent possibles incidències.
<u>Científic de dades</u>	Federat	– Col·labora en la identificació de noves oportunitats i casos d'ús de valor per al departament. – Realitza anàlisis exploratòries de dades. – Prepara, entrena i valida models analítics, aplicant tècniques de ciència de dades i seleccionant característiques rellevants. – Implementa i manté les eines per al desenvolupament de models. – Avalua la viabilitat i selecciona el model final. – Participa en l'avaluació i validació contínua del model, configurant indicadors clau de rendiment (KPI en anglès) per a l'auto-reentrenament quan sigui necessari.
<u>Enginyer d'aprenentatge automàtic</u>	Federat	– Se centra en la implementació i desplegament de sistemes d'IA. – Automatitza fases de preparació i adquisició de dades. – Optimitza els processos involucrats en la fase de desenvolupament d'IA.



A dia d'avui, es pot donar la possibilitat que alguns rols tècnics no estiguin presents en certs departaments de l'Administració de la Generalitat, mentre que en d'altres sí que poden estar-ho. Per a aquells departaments on aquests rols no estiguin definits, se suplirà aquesta mancança mitjançant l'assumpció de responsabilitats per part d'altres rols presents en el departament, o mitjançant contractació externa de serveis.

3

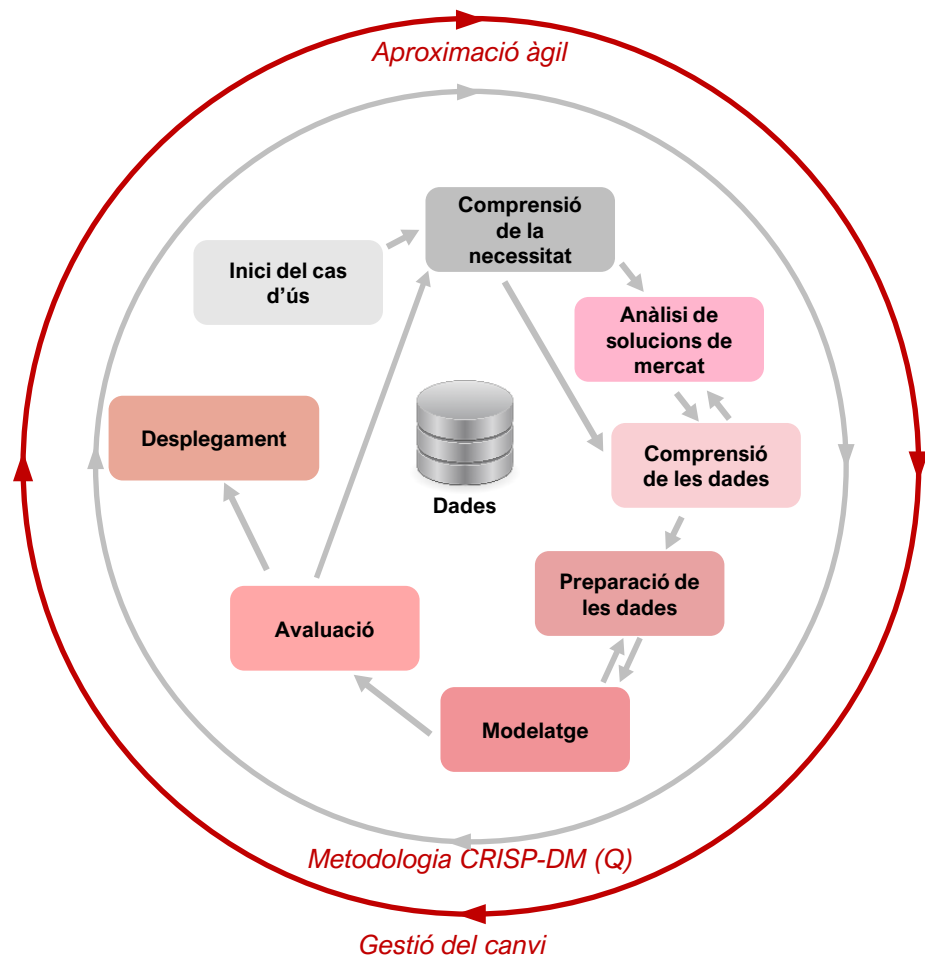
Cicle de vida – Visió Global

3. Cicle de Vida Execució

Visió global

Per al **desenvolupament dels casos d'ús d'analítica avançada o IA**, s'utilitzarà un model que híbrid de **MLOps, CRISP-DM i Àgil**. Aquest model és **CRISP-ML(Q)***, que permetrà garantir el **govern, gestió i monitoratge** de les **solucions d'analítica avançada i IA a la STA** amb uns **criteris de qualitat de forma segura**, sempre garantint l'**alineament amb els diferents models de govern definits o en definició**.

Monitoreig de costos i rendiment tècnic



Bases pel desenvolupament de casos d'ús

CRISP-ML(Q)*

Metodologia oberta que inicia l'anàlisi des d'una **perspectiva global** i posa èmfasi en el **coneixement del negoci** abans d'entrar en la manipulació de dades i la construcció.

Sostenibilitat

La metodologia **Green IT** per al càlcul de l'**impacte mediambiental** està **integrada** en les fases del cicle de vida del cas d'ús i es garanteix així el **desenvolupament** de casos d'ús **sostenibles** perquè es quantifica i considera la seva petjada de carboni

Excel·lència

L'**anàlisi i comparació de les solucions** de mercat garanteix la tria de la solució que **millor s'adapti** a les necessitats de les unitats dels Departaments i s'assegura, en tot cas, els màxims estàndards d'**ètica**, de **qualitat** i de **seguretat**.



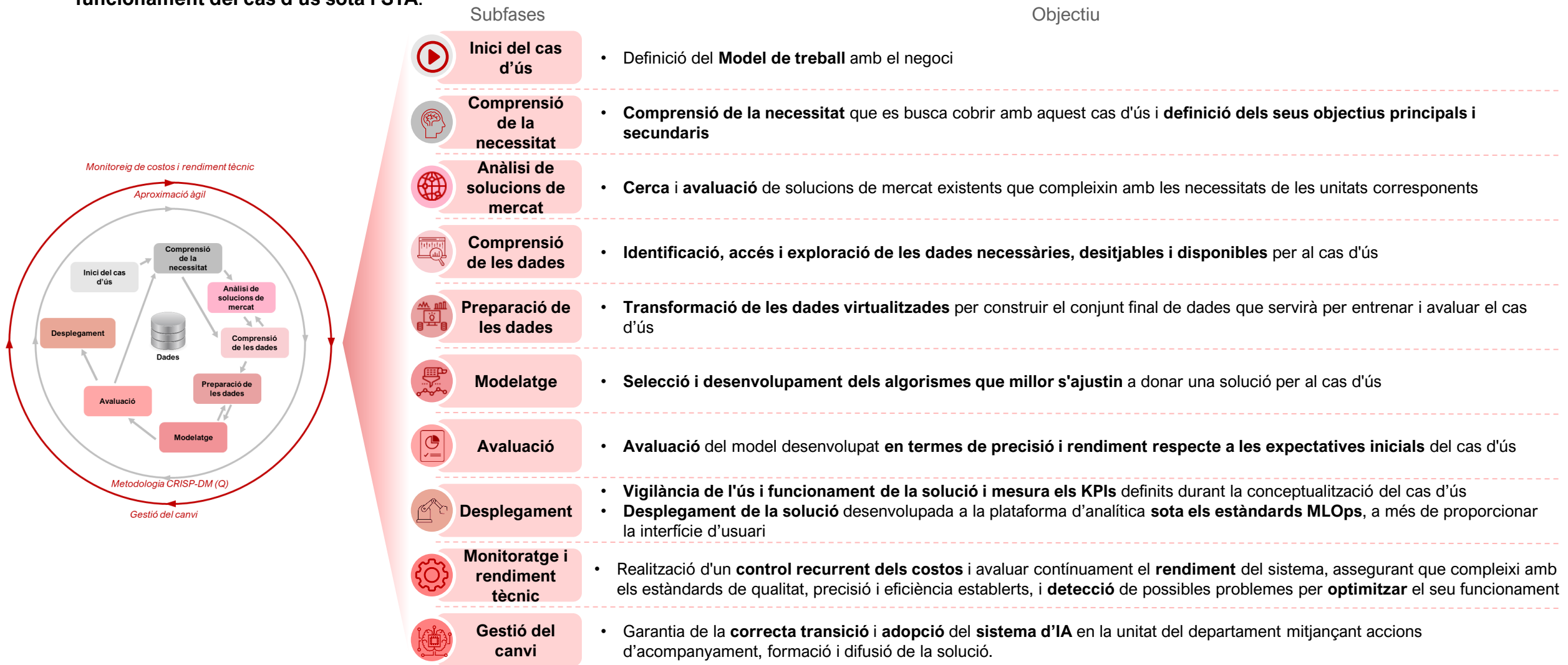
Es troben marcats amb aquesta icona les etapes del cicle de vida dels casos d'ús on s'apliqui la metodologia **Green IT** (detallada al document **2404 Metodologia Green IT v01.pptx**)

*En anglès: Cross-Industry Standard Process for the development of Machine Learning applications with Quality assurance methodology

3. Cicle de Vida Execució

Visió global

Així doncs, del marc metodològic CRISP-DM(Q) es desprenen les **fases** següents a dur a **terme des de la concepció fins a la implementació i posada en funcionament del cas d'ús sota l'STA**:



4

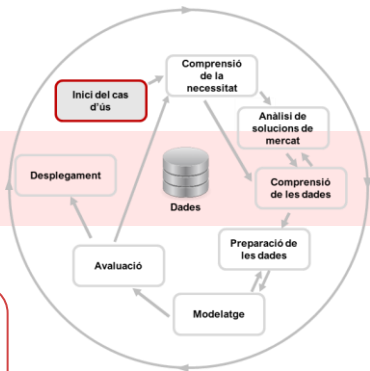
Cicle de vida – Detall

4. Cicle de Vida Execució

Detall

Inici del cas d'ús

Definició del model de treball amb el negoci



Responsable de la unitat	Es detecta una necessitat dins de la unitat en relació amb l'analítica avançada o IA. Es comunica als actors clau, com l'àrea TIC, i òrgans centrals quina és aquesta necessitat i la solució proposada.
Referent de Govern de les dades i IA	Proporcionar al responsable de la unitat la metodologia que ha de seguir per desenvolupar el cas d'ús.
Referent de Govern de les dades i IA	Concreció de l'aplicació de la metodologia (<i>Àgil</i> o <i>cascada</i> en funció dels criteris definits per CTTI) per al cas d'ús i establiment, entre d'altres, dels rols i model de relació amb la unitat del Departament.
Responsable de la unitat	Realització de la reunió d'arrencada del cas d'ús.

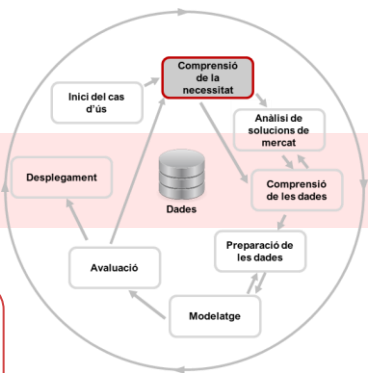
*Consultar Matriu **RACI** per a mes detall

4. Cicle de Vida Execució

Detall

Comprensió de la necessitat (1/2)

Comprensió de la necessitat que es busca cobrir amb aquest cas d'ús i definir els seus objectius principals i secundaris



Responsable de la unitat	Recull de documentació, si n'existeix.
Cap de projecte d'IA/AA	Entrevistes i dinàmiques amb agents clau per entendre la necessitat associada al cas d'ús i els requisits funcionals.
Enginyer de dades	Identificació de les dades necessàries, desitjables i disponibles, així com anàlisi de la seva qualitat i la necessitat o no de ser dades governades.
Responsable de la unitat	Revisió del tractament de dades personals (si aplica) omplint el formulari: Ampliació informació actv. tractament

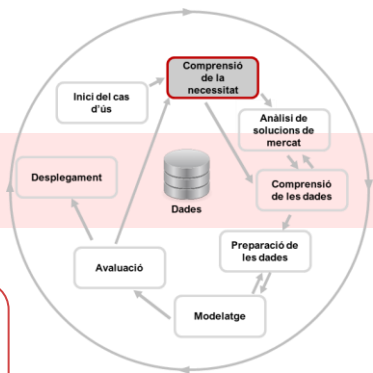
*Consultar Matriu **RACI** per a mes detall

4. Cicle de Vida Execució

Detall

Comprensió de la necessitat (2/2)

Comprensió de la necessitat que es vol cobrir amb el cas d'ús i definir-ne els objectius principals i secundaris



Responsable de la unitat	Identificació de possibles dependències i necessitats dependents d'altres Departaments i unitats.
Responsable de la unitat	Definició d'objectius i l'abast del cas d'ús (producte mínim viable i evolutius), de tal manera que faciliti la definició del ROI.
Responsable del la unitat	Aterratge del Marc de gestió de riscos al cas d'ús. Procés iteratiu al llarg de tot el cicle de vida del cas d'ús. *Per a més detall, consultar el document Marc de gestió de riscos d'Intel·ligència Artificial de l'Administració de la Generalitat de Catalunya.
Responsable de la unitat	Alta de la solució en el Registre de Sistemes d'IA i altres algorismes, i emplenar els camps amb la informació disponible. Procés iteratiu al llarg de tot el cicle de vida del cas d'ús. *Per a més detall, consultar el document Registre de sistemes d'Intel·ligència Artificial i altres algorismes.



- **Definició** del problema: context i descripció, objectius i beneficis, fonts d'informació a tenir en compte (rellevant), agents involucrats (amb els seus rols)

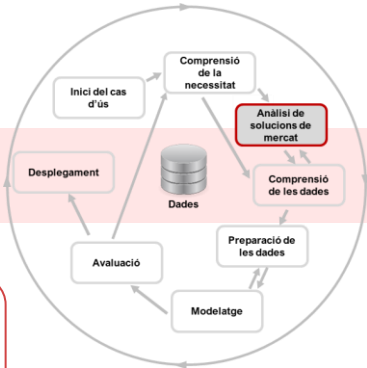
*Consultar Matriu **RACI** per a mes detall

4. Cicle de Vida Execució

Detall

Anàlisi de solucions de mercat (1/2)

Cerca i avaluació de solucions de mercat existents que compleixin amb les necessitats de les unitats corresponents



Cap de projecte d'IA/AA	Identificació de possibles solucions de mercat (<i>benchmarking</i>).
Cap de projecte d'IA/AA	Selecció d'aquelles solucions que cal avaluar, a partir d'unes variables definides.
Cap de projecte d'IA/AA	Demanda d'informació tècnica (inclòs, si escau, la relativa al FLOPs), funcional i econòmica de les solucions que permeti la seva avaluació i comparació.

*Consultar Matriu **RACI** per a mes detall

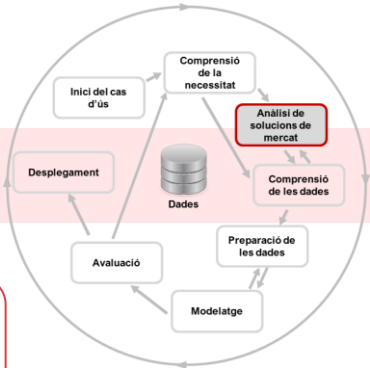
4. Cicle de Vida Execució

Detall



Anàlisi de solucions de mercat (2/2)

Cerca i avaluació de solucions de mercat existents que compleixin amb les necessitats de les unitats corresponents



Cap de projecte
d'IA/AA

Avaluació de les solucions seleccionades. Es prendrà en consideració criteris mediambientals en l'avaluació i decisió.

Cap de projecte
d'IA/AA

Contacte amb el proveïdor, realització de la demostració i avaluació de la proposta.

Cap de projecte
d'IA/AA

Selecció de la solució.

*Consultar Matriu **RACI** per a
mes detall

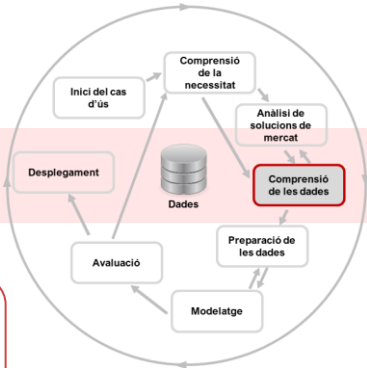
4. Cicle de Vida Execució

Detall



Comprensió de les dades (1/2)

Identificació, accés i exploració de les dades necessàries, desitjables i disponibles per al cas d'ús



Arquitecte de dades	Elaboració del document d'arquitectura/ disseny funcional i Pla de qualitat
Enginyer de dades	Disponibilització de les dades i extracció d'una mostra de les dades disponibles (obertes i/o internes)
Científic de dades	Comprensió profunda de cada variable

*Consultar Matriu **RACI** per a mes detall

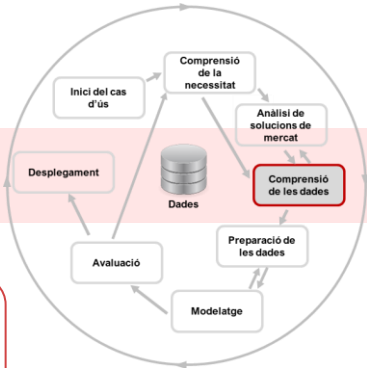
4. Cicle de Vida Execució

Detall



Comprensió de les dades (2/2)

Identificació, accés i exploració de les dades necessàries, desitjables i disponibles per al cas d'ús



Científic de dades

Anàlisis bàsiques que mostrin dispersions i correlacions entre variables

Científic de dades

Identificació de tècniques analítiques amb les quals poder abordar el cas d'ús

Científic de dades

Compartició de les principals conclusions de la caracterització de les dades als futurs usuaris funcionals del sistema resultant del cas d'ús



- **Document de recopilació de dades:** descripció detallada de les dades, contingut de les dades, relacions entre les dades, rols sobre les dades, ubicació de les dades, accessibilitat de les dades

*Consultar Matriu **RACI** per a mes detall

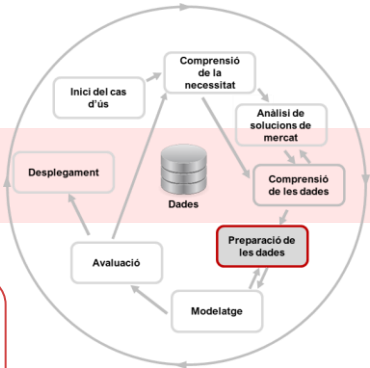
4. Cicle de Vida Execució

Detall



Preparació de les dades

Transformació de les dades virtualitzades per construir el conjunt final de dades que servirà per entrenar i avaluar el cas d'ús



Responsable de la unitat

Si cal governar les dades, en paral·lel i de forma coordinada, es realitzarà el procés per a la seva adequació a la PTD (si aplica) i governança.

Científic de dades

Reordenació i adequació de les dades.

Científic de dades

Transformació i creació de variables per al modelatge.



- **Document de preparació de les dades:** informe sobre el procés de neteja i preprocessament de les dades, incloent-hi l'eliminació de valors atípics i la normalització de variable

*Consultar Matriu **RACI** per a mes detall

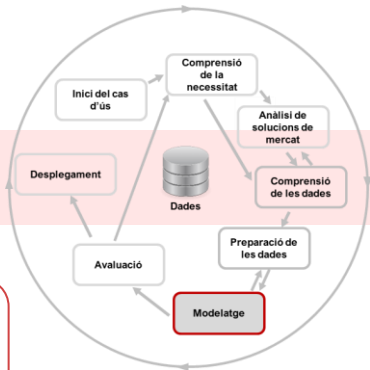
4. Cicle de Vida Execució

Detall



Modelatge (1/2)*

Selecció i desenvolupament dels algorismes que millor s'ajustin a donar una solució per al cas d'ús



*Si es tracta d'una solució de mercat, s'obvia aquesta fase.

Científic de dades	Comparativa (<i>benchmarking</i>) de tècniques que s'adaptin als requisits del cas.
Científic de dades	Selecció de l'algorisme més adequat per al cas d'ús. Es prendrà en consideració criteris mediambientals (petjada de carboni) en la decisió, detallat al document 2404 Metodologia Green IT v01.pptx
Científic de dades	Programació i parametrització òptima de la tècnica seleccionada.
Científic de dades	Mesurament de FLOPs a l'entrenament.

*Consultar Matriu **RACI** per a mes detall

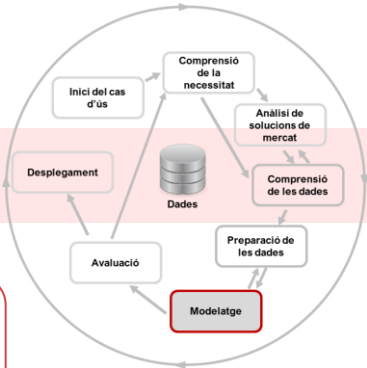
4. Cicle de Vida Execució

Detall



Modelatge (2/2)*

Selecció i desenvolupament dels algorismes que millor s'ajustin a donar una solució per al cas d'ús.



*Si es tracta d'una solució de mercat, s'obvia aquesta fase.

Científic de dades

Si supera els 10^25 FLOPs preparació de la documentació detallada de l'entrenament del model funcional requerida pel RIA.

Responsable de ciberseguretat

Validació del sistema des de la perspectiva de la ciberseguretat.

Científic de dades

Si no és validat, s'apliquen les mesures correctives necessàries.



- **Document d'entrenament del models:** descripció del procés d'entrenament del model, incloent-hi la divisió de dades en conjunts d'entrenament, validació i prova

*Consultar Matriu **RACI** per a mes detall

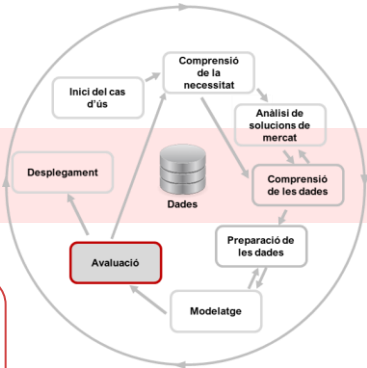
4. Cicle de Vida Execució

Detall



Avaluació (1/2)

Avaluació del model desenvolupat en termes de precisió i rendiment respecte a les expectatives inicials del cas d'ús



Responsable de la unitat	Avaluació d'impacte de protecció de dades. Assegurar que es compleixin tots els requisits de protecció de dades.
Científic de dades	Definició del model d'avaluació més adequat per al cas d'ús, d'acord amb la naturalesa de l'algorisme i les dades de testatge disponibles i prenent en consideració l'impacte mediambiental, detallat al document 2404 Metodologia Green IT v01.pptx .
Científic de dades	Implementació del model d'avaluació definit (programació de l'avaluació, testatge funcional...) i càlcul de l'impacte mediambiental.
Científic de dades	Anàlisi dels resultats d'avaluació obtinguts, no només en termes de precisió sinó també de rendiment i impacte mediambiental, i identificació d'idees que poden millorar l'algorisme.

*Consultar Matriu **RACI** per a mes detall

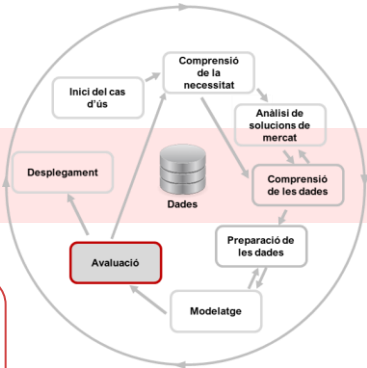
4. Cicle de Vida Execució

Detall



Avaluació (2/2)

Avaluació del model desenvolupat en termes de precisió i rendiment respecte a les expectatives inicials del cas d'ús



Científic de dades

Si escau, tornar a la fase de modelatge per ajustar paràmetres de l'algorisme o a la fase de comprensió de dades per comprovar si un altre dels algorismes pot millorar els resultats.

Responsable de la unitat

Comprovació que s'ha completat l'anàlisi ètica del cas d'ús i validar-lo èticament.

Cap de projecte d'IA/AA

Avaluació final de l'aptitud del sistema d'IA/algorisme i aprovació de la seva posada en producció.
Comunicació dels resultats de l'avaluació obtinguts al responsable de la unitat.



- **Document d'avaluació dels models:** resultats de l'avaluació del model utilitzant mètriques de rendiment, anàlisi dels resultats i conclusions sobre la qualitat i rendiment del model

*Consultar Matriu **RACI** per a mes detall

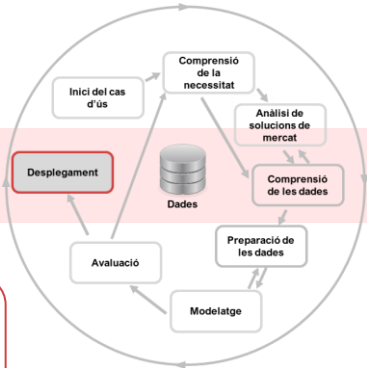
4. Cicle de Vida Execució

Detall



Desplegament

Vigilància de l'ús i funcionament de la solució i mesurar els indicadors definits durant la conceptualització del cas d'ús
Desplegament de la solució desenvolupada a la plataforma d'analítica sota els estàndards MLOps i facilitació de la interfície d'usuari



Enginyer d'aprenentatge automàtic	Automatització dels elements relacionats amb la gestió de la dada (integració, el govern de la seva qualitat i aspectes relacionats amb la seguretat).
Enginyer d'aprenentatge automàtic	Automatització d'aspectes relacionats amb el modelatge, l'entrenament i la validació, assegurant adequada traçabilitat i reproductibilitat.
Responsable del Registre	Validació que tota la informació del sistema està completa en el Registre.
Responsable de la unitat	Sessió final per donar visió del tot el procés fet i la importància del govern i la qualitat de les dades.

*Consultar Matriu **RACI** per a mes detall

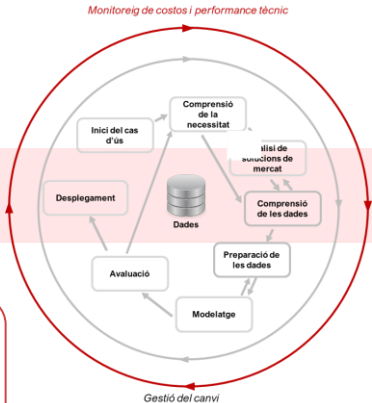
4. Cicle de Vida Execució

Detall



Monitoratge i rendiment tècnic

Control dels costos i avaluar contínuament el rendiment del sistema, assegurant que compleixi amb els estàndards de qualitat, precisió i eficiència establerts, i detectar possibles problemes per optimitzar el seu funcionament



Científic de dades	Definició i implementació de indicadors claus de rendiment.
Científic de dades	Mesurament i anàlisi periòdic dels indicadors (tasca iterativa al llarg del cicle de vida del funcionament del sistema).
Científic de dades	Identificació de possibles desviacions o indicatius que requereixin una revisió del model. (tasca iterativa al llarg del cicle de vida del funcionament del sistema)
Responsable de la unitat	Comunicacions periòdiques de l'estat de situació, el rendiment del sistema (tasca iterativa al llarg del cicle de vida del funcionament del sistema)
Responsable de la unitat	Revisió dels camps del Registre, realitzar modificacions, si s'escau.

*Consultar Matriu **RACI** per a mes detall

4. Cicle de Vida Execució

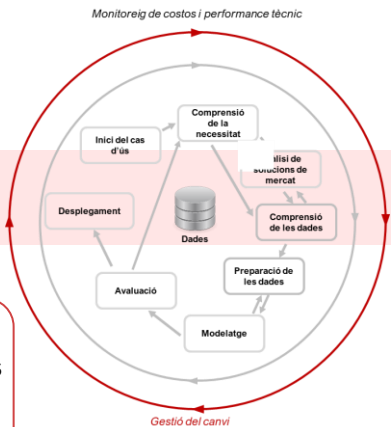
Detall



Gestió del canvi (1/2)

Garantia de la correcta transició i adopció del sistema d'IA en la unitat del departament mitjançant accions d'acompanyament, formació i difusió de la solució

Responsable de la unitat	Creació d'una estratègia de comunicació bidireccional dels objectius, beneficis i funcionament del nou model capaç d'informar els agents implicats i resoldre els seus dubtes, així com promocionar el ROI del sistema d'IA.
Responsable de la unitat	Capacitar els agents implicats mitjançant sessions de formació adaptades als diferents nivells de coneixement tècnic.
Referent de Govern de les dades i IA	Designar mentors dins les unitats implicats que pugui donar suport i orientar de forma específica els empleats públics i, a més, aconseguir un retorn del procés amb la identificació i abordatge de reptes o preocupacions.



*Consultar Matriu **RACI** per a mes detall

4. Cicle de Vida Execució

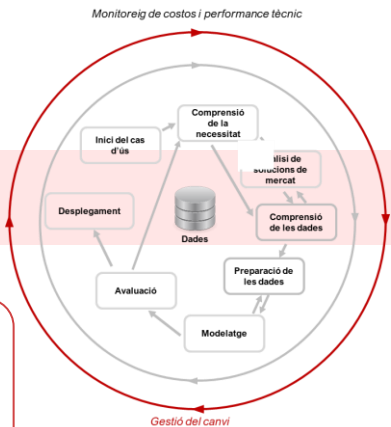
Detall



Gestió del canvi (2/2)

Garantia de la correcta transició i adopció del sistema d'IA en la unitat del departament mitjançant accions d'acompanyament, formació i difusió de la solució

Responsable de la unitat	Implicació en la gestió i seguiment del sistema d'IA o AA a l'ÀTIC corresponent per facilitar la transició de responsabilitats un cop acabat el servei.
Referent de Govern de les dades i IA	Implementació, seguiment i millora contínua de l'estratègia de comunicació i capacitació definida i del programa de mentors.
Responsable del Registre	Registre: Promoure el canvi cultural mitjançant activitats de difusió i formació per a tots els perfils usuaris.



*Consultar Matriu **RACI** per a mes detall



Generalitat de Catalunya

Centre de Telecomunicacions i Tecnologies de la Informació (CTTI)

ctti.gencat.cat