Medical Coding @Bayer

Replacing IBM Watson

13 November 2020

RESTRICTED

Agenda

- Summary
- Current Situation and Goals
- Project Overview
- Project Plan
- Business Case
- Team Overview and Required Profiles
- Next Steps

Summary

Project Overview

We defined a clear objective for the project based on the identified scope and derived cost and schedule estimates

Objective of the project

The project objective is to replace IBM Watson for the 1st tier medical coding by developing and integrating state-of-the-art NLP models with approximately similar accuracy that ensures a smooth replacement by March 2022. The solution should furthermore be sustainable in terms of more use cases such as processing real world evidence which are envisioned to be included by Bayer further down the road.

Scope

Coding Algorithm Development

- Develop Medical Coding algorithm for **C**linical **T**rials (**M**edical **H**istory& **A**dverse **E**vents) data
- **Not in Scope:**
  - WhoDD Coding
  - Pharmacovigilance (MH&AE) Coding

Architecture & Interfaces

- Design and build Architecture and Pipeline for a Cloud-based Machine Learning solution
- Define and build Interfaces to current Medical Coding platform (**M**atch**P**oint **C**oder)

Testing & Deployment

- User Testing with Medical Coding Team
- GxP Validation (Change request to MPC)
- Deployment of developed algorithms into existing IT infrastructure and Medical Coding processes
- (Complete) Decommission of IBM Watson

Estimated Schedule

- Kick-off: January 2021
- Phase 1: January 2021 – March 2021 (2.5 months)
- Phase 2: March 2021 – December 2021 (9.5 months)
- Phase 3: January 2022 – May 2022 (5 months)
- End of Transition Phase: May 2022 (17 months total)

Cost Estimate

• Phase 1:	200.000 €	(approval pending)
• Phase 2:	1.121.700 €	
• Phase 3:	263.700 €	
• **Total:**	**1.585.400 €**	

Goals & Benefits

Watson’s replacement offers great opportunities which are also in line with Bayer’s superior strategy

Reduce run cost of medical coding software by avoiding high annual license fees

Utilize most recent advances in NLP domain to **increase** accuracy for medical coding

Take advantage of available in-house skills, i.e. NLP Experts and Medical Coder

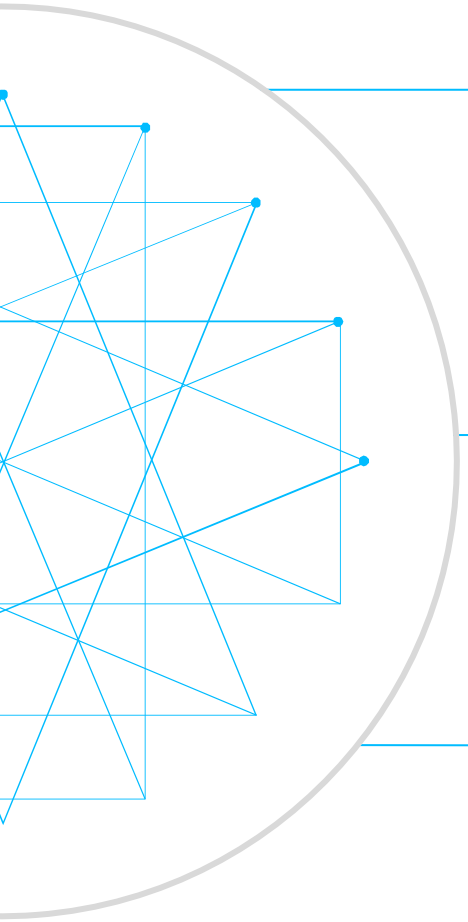
Enable future use cases such as medical coding of Real World Data

Ensure available Business IP stays within Bayer and **develop** new IP, i.e. capitalization of algorithms and/or patents

The goals and benefits that replacing IBM Watson aims at are aligned with and contribute to key strategic programs. This includes the central coding approach where Bayer is unique as well as the Digital Roadmap.

Important Parameters

There are certain dates and details that determine the project timeline



IBM Watson Contract

The contract with IBM expires on 31 March 2022. In case it should be renewed, a Purchas Order must be placed at least 90 days prior to expiration. This requires internal decision processes and contract negotiations with IBM and therefore it is crucial to consider this with regard to the project timeline.

Budget impact

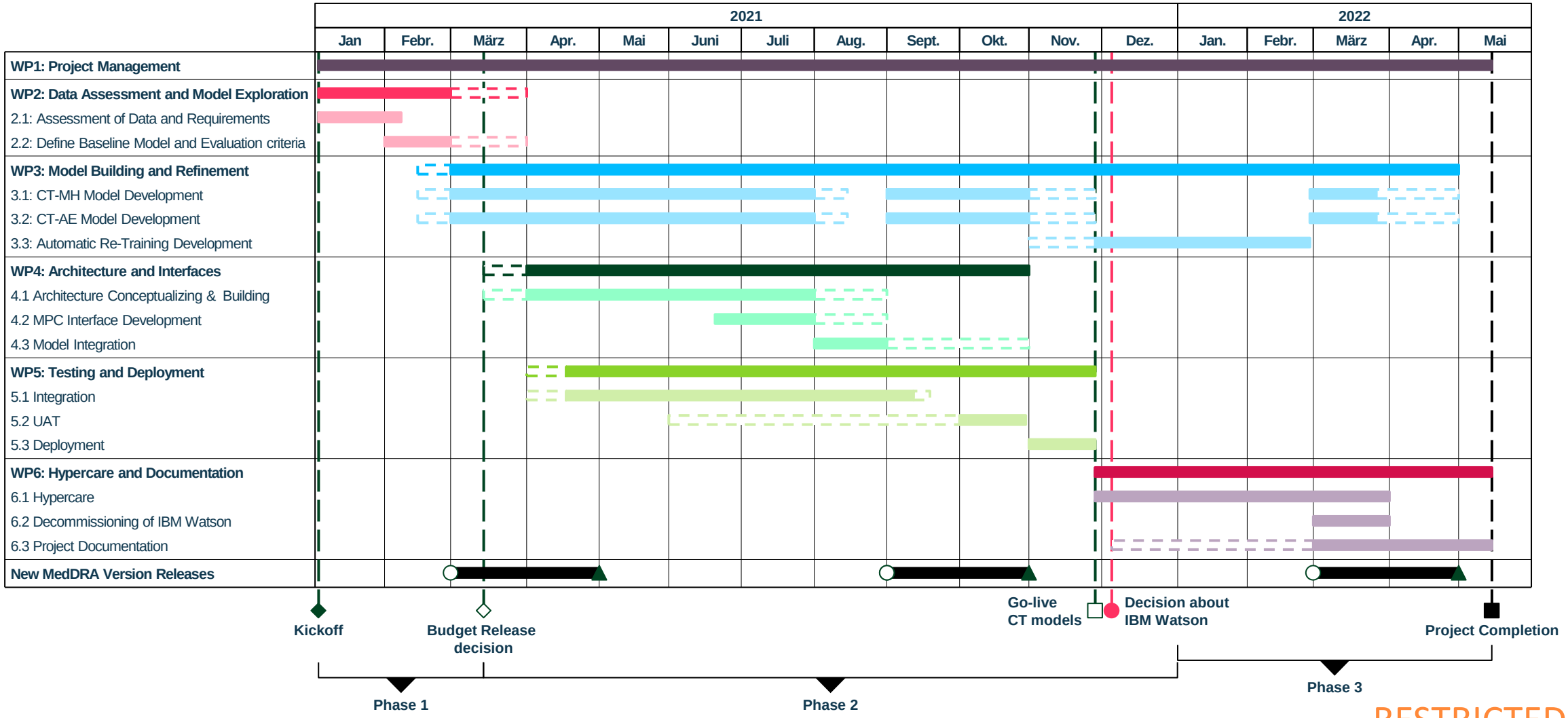
Due to the current situation at Bayer, budget considerations for 2021 are ongoing. However, the complexity of the task and the expiration of Watson's contract mandate a start of activities in January 2021. This ensures enough time to avoid further license fees to IBM as of 2022.

MedDRA Updates

The MedDRA maintenance organization releases new versions of the dictionary twice a year, on 1 March and 1 September. The required changes need to be implemented on the first Monday of May and November, respectively. This hast to be considered when planning model development and deployment.

Project Plan Draft

3 Phases | 17 Months of Activities | Go-Live in November 2021



Project Overview - Work Packages

There are six major Work Packages with specific outcomes and effort estimates

	Short Description	Outcome	Effort and Cost Estimate
WP1 Project Management	• Manage project and activities • Coordinate available resources • Ensure compliance with internal processes	• Enablement and facilitation of team activities • Project documentation and communication of results • Risks are continuously monitored and can be mitigated in time	• PD: 346 • Costs: 345.400 €
WP2 Data Assessment & Model Exploration	• Develop understanding for data and requirements • Assess previous work • Define baseline and metrics	• Relevant data available • Baseline model and evaluation criteria are defined • Promising approaches for Model development are evaluated	• PD: 152 • Costs: 119.900 €
WP3 Model Building & Refinement	• Building AI models for medical coding • Develop automatic update solution for new MedDRA versions	• AI Models for different Clinical Trials terms (MH&AE) coding • Automatic integration of MedDRA updates	• PD: 634 • Costs: 550.800 €
WP4 Architecture & Interfaces	• Design and build cloud environment including necessary interfaces • Documentation according to Bayer requirements	• Cloud based infrastructure and interfaces with MPC • GxP compliant documentation and validation	• PD: 252 • Costs: 245.400 €
WP5 Testing & Deployment	• Testing and validation of new system • Integration and Deployment	• Successfully performed UAT as well as integration and validation testing • Successful Deployment and Go-Live • Complete set of validation documents	• PD: 375 • Costs: 276.400 €
WP6 Hypercare & Documentation	• Disconnecting IBM Watson from the coding environment • Monitoring the new system • Close the project	• IBM Watson successfully decommissioned • Stable new system • Full project documentation	• PD: 65 • Costs: 54.500 €

Project Overview – Guiding Principles

Key points that determine the development process

CHANGE DIMENSIONS

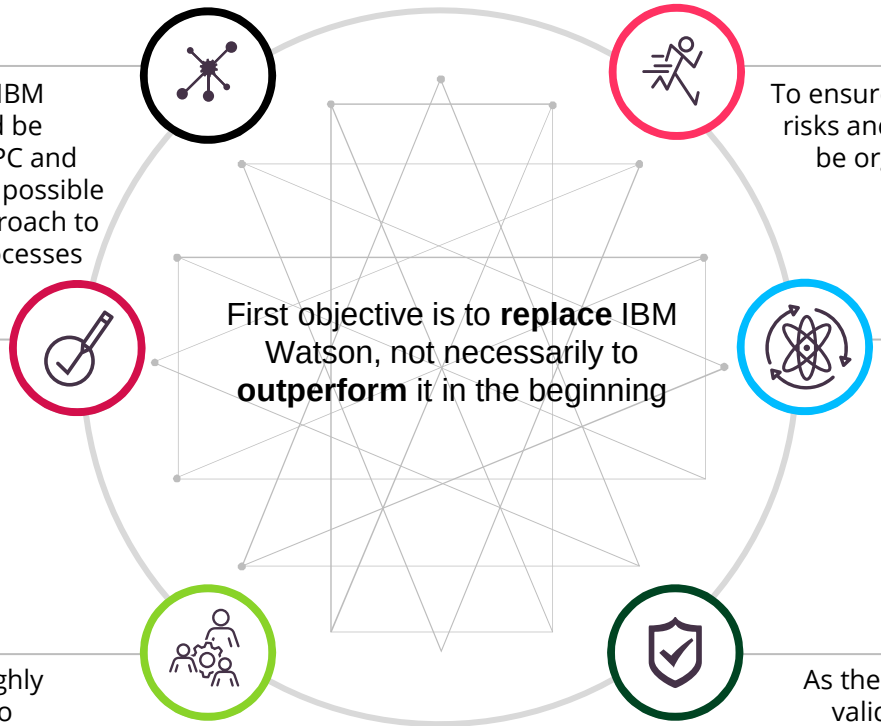
As the MatchPoint Coder currently interacts with IBM Watson via an interface, no major changes should be made to the MPC itself. The interface between MPC and the new system should therefore be as similar as possible to the old one to ensure a “minimal-invasive” approach to MPC. There will be no change to any business processes due to this project.

REQUIREMENTS

The requirements for the new system are based on the original IBM Watson ones. However, they were adjusted due to reduced scope and minor additional features. They build the foundation for the development phase and need to be aligned on with all team members.

STAKEHOLDER INVOLVEMENT

Medical Coding is a complex task that requires highly skilled human resources. For a computer model to perform well on such a task, it needs to incorporate the knowledge, especially when it comes to specific differences, e.g. between CT & AE term types. Therefore, it is essential to involve business stakeholders in evaluation and feedback loops.



WAYS OF WORK

To ensure flexibility and minimize reaction times to emerging risks and issues, the development of the new system should be organized in two- or three-week sprints. Those include sprint planning and review sessions with selective participation of SMEs and Stakeholders.

METHODS

The ambition and expectations with regard to the new system is that it uses state-of-the-art AI methods wherever appropriate. This should contribute to Bayer’s Digital Roadmap as well as to the vision to include future Use Cases such as coding real world evidence, thus providing a sustainable coding engine.

QUALITY CONTROL

As the project takes place in an GxP-relevant environment, validation and process compliance hast to be monitored and ensured constantly.

Major Risks and Dependencies

Emerging and present risks and dependencies need to be considered, anticipated and mitigated – starting now

Risks

Budget Risk

The pending 200.000€ budget for 2021 is estimated to last for the first 2.5 months of project activity. Additional budget is therefore required and approval of such should be prioritized in order to avoid any delays and ensure project continuity.

Time Risk

IBM Watson's contract expiration sets a hard deadline for developing and implementing a new coding solution. Even a one year extension should be avoided due to additional cost. Hence, any delays with regard to the outlined schedule pose an imminent risk for the overall project success and should be avoided.

Quality Risk

Developing and implementing state-of-the-art machine learning methods is always subject to certain risks of failure. Most recent research results that seem promising at first can turn out to not be performing well on the actual available data and vice versa. The chosen approach should therefore continuously be evaluated and adjusted, if necessary.

Dependencies

Available Resources and Staffing

Due to the tight schedule and the project's complexity, many resources from different departments (Business, IT, external) need to be involved, during peak times at full capacity. Hence, it is essential to already communicate capacity needs and initiate a transparent staffing process.

Access to relevant data

The performance of the new coding model is highly dependent on the available training data. This includes AutoCoder inputs- and outputs as well as final results of the coding process. Also data from intermediate steps, such as Watson's five proposals or initiated queries will be required.

Furthermore, additional data such as Clinical Trials full data or electronic health records might be required for language model fine-tuning. This might oblige additional data preparation steps which need to be anticipated adequately.

Business Case

The cost and effort estimates result in a positive Business Case, further supported by qualitative benefits

Quantitative

Cost	2021	2022	2023	2024	2025	2026	Total
One time cost	1.321.700 €	263.700 €					**1.585.400 €**
Run cost		60.000 €	60.000 €	60.000 €	60.000 €	60.000 €	**300.000 €**
Total*	**1.321.700 €**	**323.700 €**	**60.000 €**	**60.000 €**	**60.000 €**	**60.000 €**	**1.885.400 €**

*Not included in estimates are one-time and run cost for cloud resources as they are controlled centrally. They are estimated with ~40.000€ per year.

Benefits	2021	2022	2023	2024	2025	2026	Total
Additional revenue	- €	- €	- €	- €	- €	- €	**- €**
Savings	- €	532.995 €	532.995 €	532.995 €	532.995 €	532.995 €	**2.664.975 €**
Cost Avoidance		79.949 €	79.949 €	79.949 €	79.949 €	79.949 €	**399.746 €**
Total							**3.064.721 €**

Pay-Back Period (in years) **3,31**

Net Present Value **489.207 €**

Qualitative

- Support Digital Roadmap through expanding Bayer's internal AI skills and expertise
- Ensure available Business IP stays within Bayer and further develop new IP, i.e. AI algorithms and/or patents
- Utilize most recent advances in NLP domain to increase accuracy for medical coding
- Enable future use cases such as medical coding of Real World Data
- Reduce dependency from external vendors to allow for further development and improvement

Next Steps

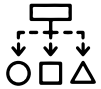
With the proposed project start being only 6 weeks away, internal initialization has to start now

Request and secure budget for 2021 and 2022 to avoid delays in project delivery

Determine available employee resources and **identify** needs for external support

Involve relevant stakeholders and **close** information gaps (e.g. regarding cloud infrastructure)

Gather relevant AutoCoder and omissions data and identify further data sources



Prepare project kick-off and set up project management

Validate User Requirements and Effort Estimates, adjust when necessary

Initiate further initiatives required for a full replacement of IBM Watson (e.g. UMC Drug Coding PoC)

Start negotiations with IBM regarding a 1-year extension to have a fallback option and sharpen Business Case

Current Situation & Goals

Current Situation – Starting Point

To meet regulatory requirements and to speed up medical coding processes, Bayer is currently using IBM Watson as an external software service.

This results in high costs due to annual fees for IBM which are likely to increase after the expiration of the initial contract

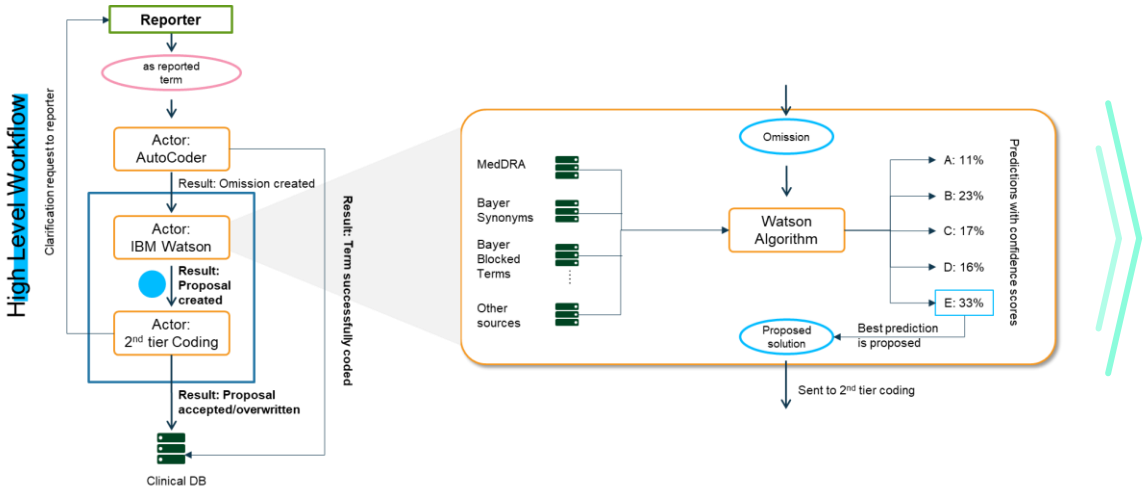
Also, Bayer has to share in-depth knowledge regarding the medical coding process and its standards with IBM on a continuous basis while no major updates have been made to the Watson coding algorithm since implementation

At the same time, Bayer AG employees are highly skilled and would allow developing and running an in-house solution, resulting in considerations about replacing IBM and developing such a solution

The goal of this project was therefore to evaluate and conceptualize the development of an in-house solution for medical coding with respect to cost, people and technology

Current Situation – Medical Coding with IBM Watson

The Medical Coding process is highly dependent on Watson, yet there are major issues coming along



The high-level medical coding workflow ensures regulatory conformity as well as consistency. While the rule-based AutoCoder is able to encode 45-50% of reported terms successfully, the omissions that are left are processed via a “Four-eyes-approach” where IBM Watson proposes solutions subject to approval through a human medical coder.

As of today, IBM Watson is used for proposing solutions to terms generated from Clinical Trials and Pharmacovigilance data based on both, the MedDRA and WhoDD dictionaries. The total volume of terms processed by Watson is about 55.000-60.000 terms per month and while the accuracy for PV data is above 90%, the CT accuracy falls well behind that with around 75%.

IBM Watson main issues

Quality

Since its full implementation in April 2017, IBM did not improve the coding algorithm, nor are they planning to do so. At the same time, Watson struggles with rather basic issues such as spelling mistakes that lead to wrong proposals. However, Bayer has no insights or access to the algorithm which can be considered as a Blackbox.

(Future) Costs

With the ending of the initial 5 years contract period on 31 March 2022, the high license fees of about 530.000€ per year are expected to increase by up to 30% resulting in a not justifiable cost-benefit balance

Current Situation – Advances in NLP domain

Since the start of Watson‘s PoC phase, NLP has come a long way

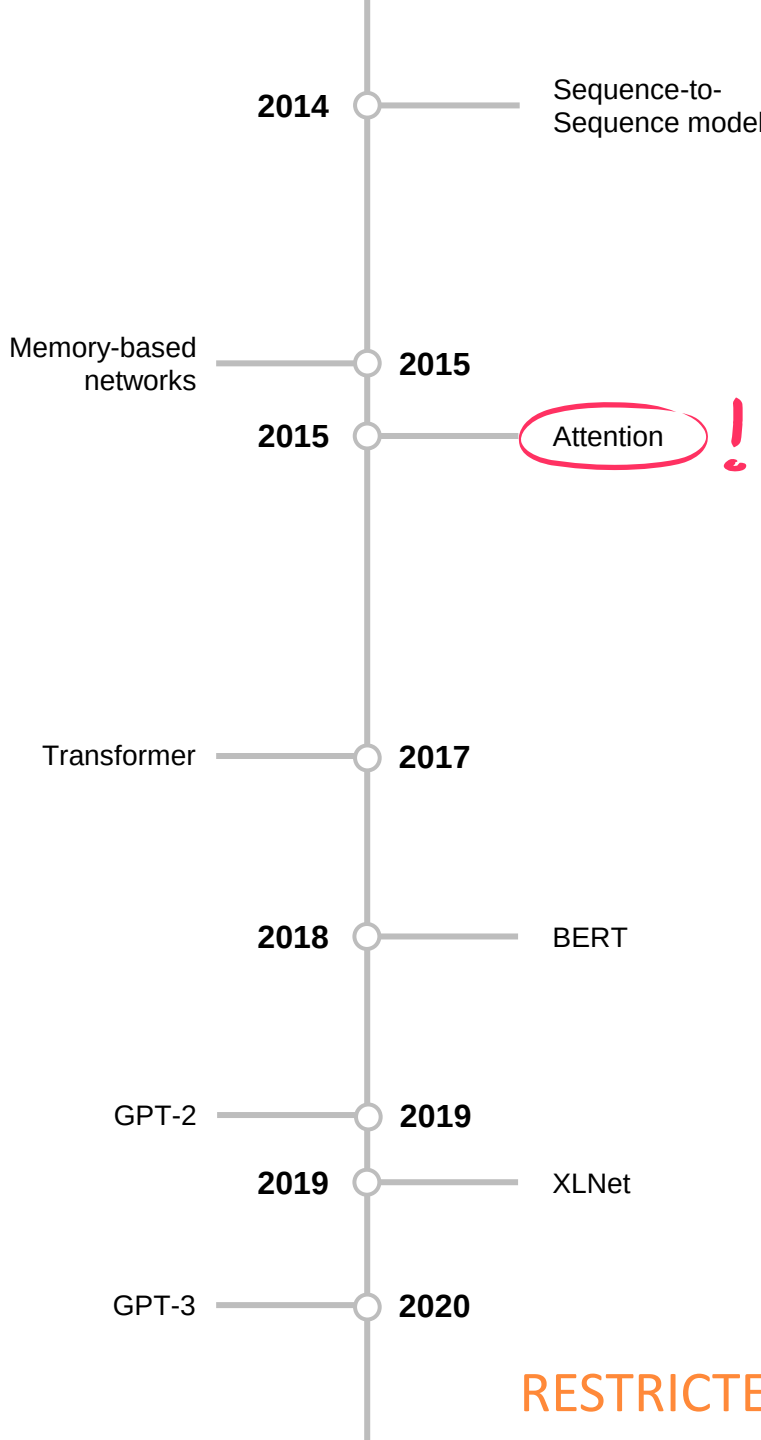
Enabled by deep learning and rapid increase in computational power, the Natural Language Processing domain witnessed unprecedented leaps over the last decade, resulting in state-of-the-art models and architectures

The introduction of the attention mechanism in 2015 was key for a broad range of new model architectures such as BERT or XLNet that outperform previous approaches on all relevant NLP tasks

Main advantage of recent language models is sensitivity to context and semantics, which allows to differentiate between two or more meanings of the same word (e.g. *Apple*, the fruit and *Apple*, the Company)

Most importantly in the context of this project, latest language models can and have been fine-tuned for usage in domains that are sensitive to specific vocabularies, such as medicine, biology or law. This results in highly accurate and domain-specific word representations which are likely to increase performance of downstream tasks such as Medical Coding.

It can be expected that developing in-house solution with state-of-the-art NLP models not only allows for better results but also enables Bayer to constantly improve model performance without being dependent on external vendors



Goals & Benefits

Watson’s replacement offers great opportunities which are also in line with Bayer’s superior strategy

Avoid low performance sites

The time to make informed decisions

The internal and external cost saving

The adoption of PoC for the new studies

Ensure available Business IP stays within Bayer and **develop** new IP, i.e. capitalization of algorithms and/or patents

The goals and benefits that replacing IBM Watson aims at are aligned with and contribute to key strategic programs. This includes the central coding approach where Bayer is unique as well as the Digital Roadmap.

Project Overview

Project Overview

We defined a clear objective for the project based on the identified scope and derived cost and schedule estimates

Objective of the project

The project objective is to replace IBM Watson for the 1st tier medical coding by developing and integrating state-of-the-art NLP models with approximately similar accuracy that ensures a smooth replacement by March 2022. The solution should furthermore be sustainable in terms of more use cases such as processing real world evidence which are envisioned to be included by Bayer further down the road.

Scope

Coding Algorithm Development

- Develop Medical Coding algorithm for **C**linical **T**rials (**M**edical **H**istory& **A**dverse **E**vents) data
- **Not in Scope:**
  - WhoDD Coding
  - Pharmacovigilance (MH&AE) Coding

Architecture & Interfaces

- Design and build Architecture and Pipeline for a Cloud-based Machine Learning solution
- Define and build Interfaces to current Medical Coding platform (**M**atch**P**oint **C**oder)

Testing & Deployment

- User Testing with Medical Coding Team
- GxP Validation (Change request to MPC)
- Deployment of developed algorithms into existing IT infrastructure and Medical Coding processes
- (Complete) Decommission of IBM Watson

Estimated Schedule

- Kick-off: January 2021
- Phase 1: January 2021 – March 2021 (2.5 months)
- Phase 2: March 2021 – December 2021 (9.5 months)
- Phase 3: January 2022 – May 2022 (5 months)
- End of Transition Phase: May 2022 (17 months total)

Cost Estimate

• Phase 1:	200.000 €	(approval pending)
• Phase 2:	1.121.700 €	
• Phase 3:	263.700 €	
• **Total:**	**1.585.400 €**	

Project Overview - Work Packages

There are six major Work Packages with specific outcomes and effort estimates

	Short Description	Outcome	Effort Estimate
WP1 Project Management	• Manage project and activities • Coordinate available resources • Ensure compliance with internal processes	• Enablement and facilitation of team activities • Project documentation and communication of results • Risks are continuously monitored and can be mitigated in time	• PD: 346 • Costs: 345.400 €
WP2 Data Assessment & Model Exploration	• Develop understanding for data and requirements • Assess previous work • Define baseline and metrics	• Relevant data available • Baseline model and evaluation criteria are defined • Promising approaches for Model development are evaluated	• PD: 152 • Costs: 119.900 €
WP3 Model Building & Refinement	• Building AI models for medical coding • Develop automatic update solution for new MedDRA versions	• AI Models for different Clinical Trials terms (MH&AE) coding • Automatic integration of MedDRA updates	• PD: 634 • Costs: 550.800 €
WP4 Architecture & Interfaces	• Design and build cloud environment including necessary interfaces • Documentation according to Bayer requirements	• Cloud based infrastructure and interfaces with MPC • GxP compliant documentation and validation	• PD: 252 • Costs: 245.400 €
WP5 Testing & Deployment	• Testing and validation of new system • Integration and Deployment	• Successfully performed UAT as well as integration and validation testing • Successful Deployment and Go-Live • Complete set of validation documents	• PD: 375 • Costs: 276.400 €
WP6 Hypercare & Documentation	• Disconnecting IBM Watson from the coding environment • Monitoring the new system • Close the project	• IBM Watson successfully decommissioned • Stable new system • Full project documentation	• PD: 65 • Costs: 54.500 €

Project Overview – Guiding Principles

Key points that determine the development process

CHANGE DIMENSIONS

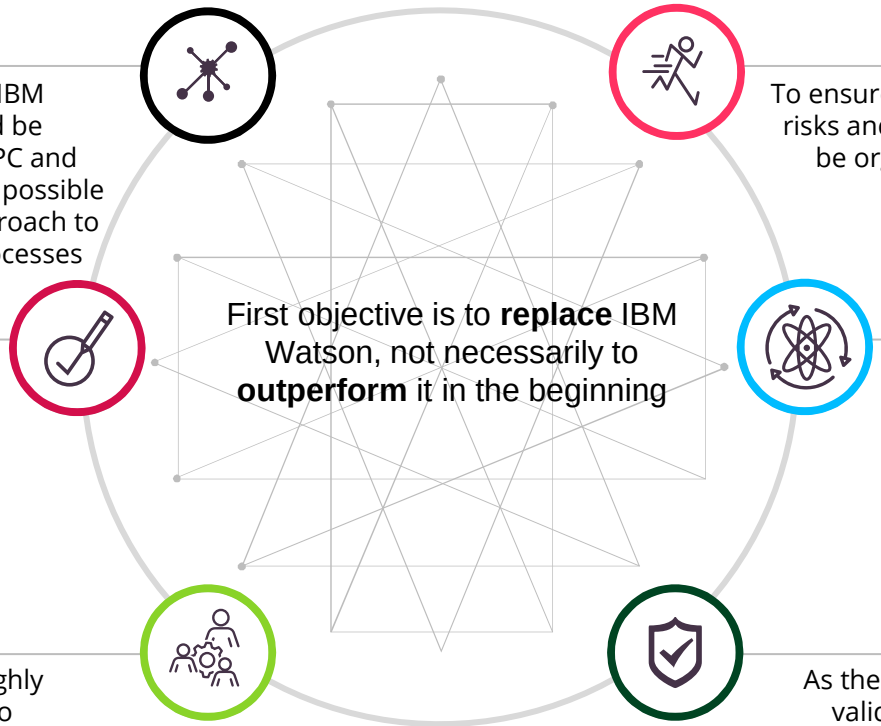
As the MatchPoint Coder currently interacts with IBM Watson via an interface, no major changes should be made to the MPC itself. The interface between MPC and the new system should therefore be as similar as possible to the old one to ensure a “minimal-invasive” approach to MPC. There will be no change to any business processes due to this project.

REQUIREMENTS

The requirements for the new system are based on the original IBM Watson ones. However, they were adjusted due to reduced scope and minor additional features. They build the foundation for the development phase and need to be aligned on with all team members.

STAKEHOLDER INVOLVEMENT

Medical Coding is a complex task that requires highly skilled human resources. For a computer model to perform well on such a task, it needs to incorporate the knowledge, especially when it comes to specific differences, e.g. between CT & AE term types. Therefore, it is essential to involve business stakeholders in evaluation and feedback loops.



WAYS OF WORK

To ensure flexibility and minimize reaction times to emerging risks and issues, the development of the new system should be organized in two- or three-week sprints. Those include sprint planning and review sessions with selective participation of SMEs and Stakeholders.

METHODS

The ambition and expectations with regard to the new system is that it uses state-of-the-art AI methods wherever appropriate. This should contribute to Bayer’s Digital Roadmap as well as to the vision to include future Use Cases such as coding real world evidence, thus providing a sustainable coding engine.

QUALITY CONTROL

As the project takes place in an GxP-relevant environment, validation and process compliance hast to be monitored and ensured constantly.

Major Risks and Dependencies

Emerging and present risks and dependencies need to be considered, anticipated and mitigated – starting now

Risks

Budget Risk

The pending 200.000€ budget for 2021 is estimated to last for the first 2.5 months of project activity. Additional budget is therefore required and approval of such should be prioritized in order to avoid any delays and ensure project continuity.

Time Risk

IBM Watson's contract expiration sets a hard deadline for developing and implementing a new coding solution. Even a one year extension should be avoided due to additional cost. Hence, any delays with regard to the outlined schedule pose an imminent risk for the overall project success and should be avoided.

Quality Risk

Developing and implementing state-of-the-art machine learning methods is always subject to certain risks of failure. Most recent research results that seem promising at first can turn out to not be performing well on the actual available data and vice versa. The chosen approach should therefore continuously be evaluated and adjusted, if necessary.

Dependencies

Available Resources and Staffing

Due to the tight schedule and the project's complexity, many resources from different departments (Business, IT, external) need to be involved, during peak times at full capacity. Hence, it is essential to already communicate capacity needs and initiate a transparent staffing process.

Access to relevant data

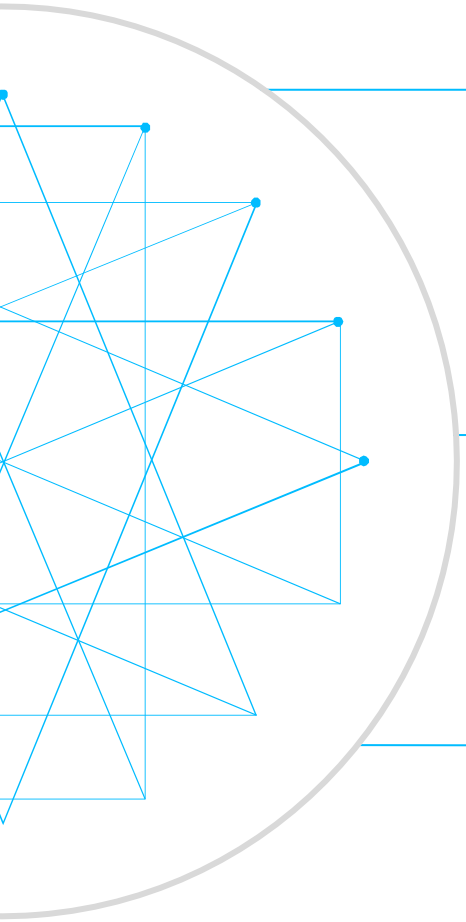
The performance of the new coding model is highly dependent on the available training data. This includes AutoCoder inputs- and outputs as well as final results of the coding process. Also data from intermediate steps, such as Watson's five proposals or initiated queries will be required.

Furthermore, additional data such as Clinical Trials full data or electronic health records might be required for language model fine-tuning. This might oblige additional data preparation steps which need to be anticipated adequately.

Project Plan

Important Parameters

There are certain dates and details that determine the project timeline



IBM Watson Contract

The contract with IBM expires on 31 March 2022. In case it should be renewed, a Purchas Order must be placed at least 90 days prior to expiration. This requires internal decision processes and contract negotiations with IBM and therefore it is crucial to consider this with regard to the project timeline.

Budget impact

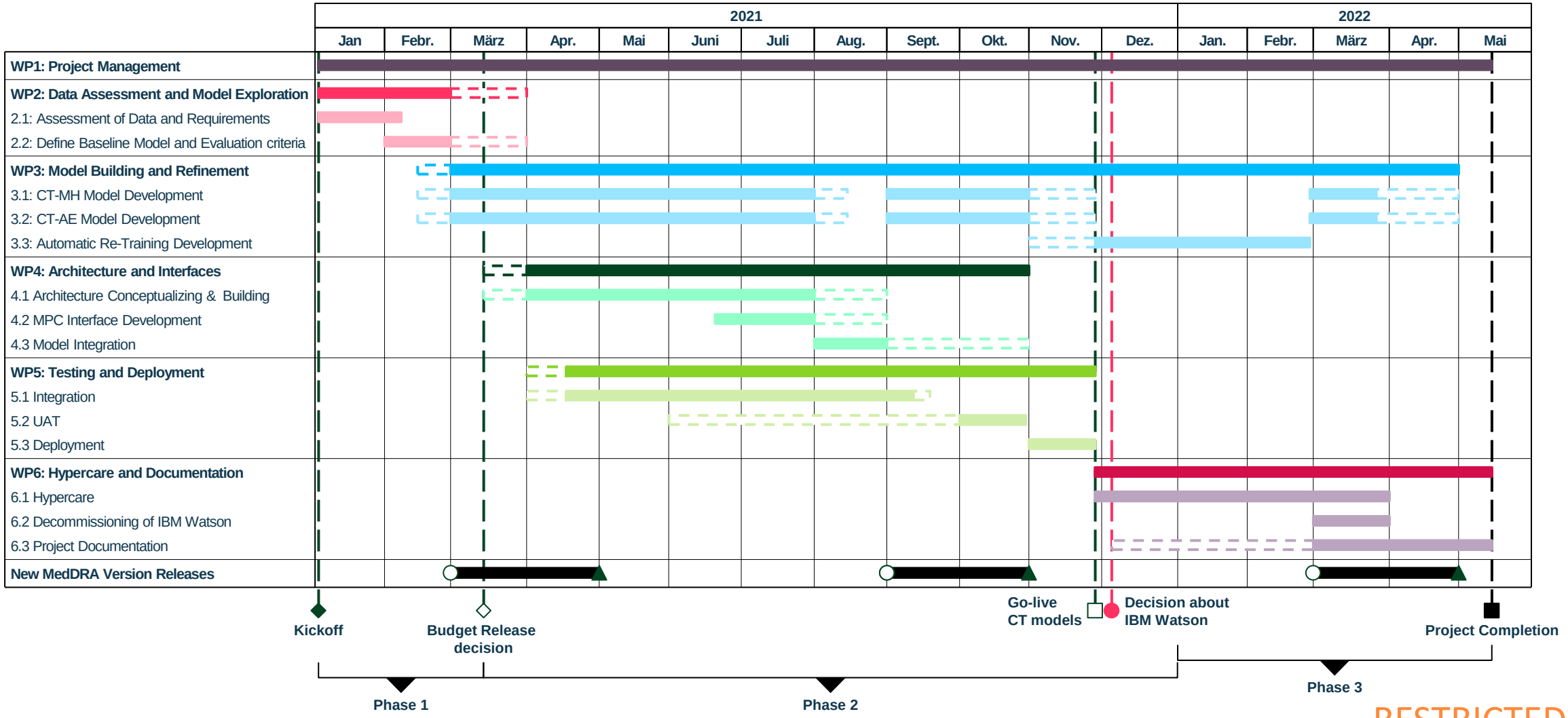
Due to the current situation at Bayer, budget considerations for 2021 are ongoing. However, the complexity of the task and the expiration of Watson's contract mandate a start of activities in January 2021. This ensures enough time to avoid further license fees to IBM as of 2022.

MedDRA Updates

The MedDRA maintenance organization releases new versions of the dictionary twice a year, on 1 March and 1 September. The required changes need to be implemented on the first Monday of May and November, respectively. This hast to be considered when planning model development and deployment.

Project Plan Draft

3 Phases | 17 Months of Activities | Go-Live in November 2021



Project Initialization

Work Package 0 | Currently ongoing

Objective

In order to initialize the project, detailed planning of time, budget and resources has to be performed as well as risk assessments and an overall project evaluation. Furthermore, important parameters such as scope and infrastructure decisions need to be defined.

Main Activities

Project Planning

- Define Scope with Business stakeholders
- Plan time, budget and resource needs and allocations
- Determine needs for Third-party vendors and initiate requests accordingly
- Perform risk assessment and mitigations
- Prepare preparation kickoff and communicate with stakeholders
- Perform project evaluation regarding strategic fit, compliance and value (tbd if still necessary)

IT & Data

- Perform initial infrastructure assessment in terms of which platform is suitable
- Address IT-Security and Data privacy issues
- Gather necessary data and information needed for project start

Involved Stakeholders

- Project Manager
- Business Stakeholder
- Data Science Team(s) (as required)
- SMEs (as required)

Outcomes

- Time, budget and resources plans are available and validated with relevant stakeholders
- Necessary parameters such as scope and infrastructure decisions are defined
- Risks are identified and can be mitigated
- Relevant data is available

Project Management

Work Package 1 | Continuous Activities

Objective

The objective of this Work Package is to manage the project and its activities, coordinate available resources and ensure compliance with internal processes to facilitate the overall project success.

Main Activities

Project Oversight

- Responsible for delivering the project in time and budget considering the scope
- Coordinating resources and timetables with stakeholders, lead workshops
- Managing project progress, assessing risks and adapt project work as required

SOP Compliance & Documentation

- Ensure project compliance and proper documentation (incl. GxP validation)

Stakeholder communications & support

- Planning and leading of Kick-off Workshop, User Testing & Deployment
- Provide team members with needed information & data
- Communicate project status to Management/Business stakeholders

Involved Stakeholders

- Project Manager

Outcomes

- Enablement and facilitation of team activities
- Project documentation and communication of results is in line with internal requirements and stakeholder expectations
- Risks are continuously monitored and can be mitigated in time

Data Assessment and Model Exploration

Work Package 2 | January – February 2021

Objective

In Work Package 2, the objective is to develop a surrounding understanding of the available data, assess previous work such as IBM Watson & PoC results and build a baseline model with specific evaluation criteria that serve as measurement for future model results.

Main Activities

Assess Data and previous work

- Gather available and relevant data
- Develop an understanding for User Requirements and data
- Review PoC Sprint results from 2019 and evaluate results
- Review IBM Watson performance

Build baseline model

- Develop initial data ingestion pipeline
- Based on model trade-off considerations and scope, build baseline model (e.g. on PoC sprint results)
- Evaluate baseline model and derive potential for “Quick wins”

Define evaluation criteria and hypotheses

- Based on IBM Watson accuracy, scope and model trade-off considerations, decide on metrics for evaluation (e.g. LLT vs PT, F1 Score vs. Accuracy, etc.)
- Define hypotheses that guide model development

Involved Stakeholders

- Data Science Team(s)
- Project Manager (as required)
- Business Stakeholder (as required)

Outcomes

- All relevant data for model training, testing and validation is available
- A baseline model as well as evaluation criteria are defined and serve as a foundation for all further models
- With respect to scope and model trade-off considerations, most promising approaches are identified for model development

Deep Dive: WP2 - Data Assessment and Model Exploration

Key questions when assessing and exploring data

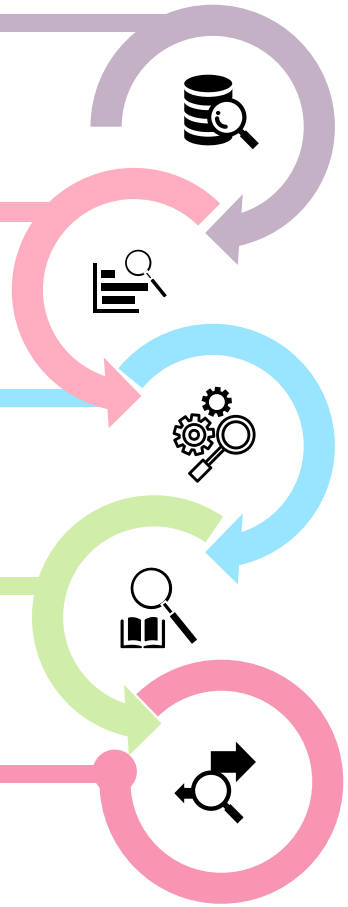
Data availability:
Is all data that is required, such as omissions dataset, IBM Watson results, dictionaries, present?
Are quality and quantity of the data sufficient?

Data distribution:
How is the data distributed over MedDRA classes?
Which are classes that are quite rare and why?
Are certain classes more important than others?

Output exploration:
What patterns can be identified within the omissions data?
Where do Watson and the AutoCoder struggle? Systematic Error(s)?
How is Watson accuracy determined (LLT vs. PT Level)?

Coding specifics & user requirements:
Which implications do the Bayer coding standards have on the data and future models?
Which user requirements can be more easily or more difficult be implemented?

Model hypotheses:
Which approach seems most promising for quick wins?
Which implications do different approaches have on additional data needs?
Which measures seem reasonable for model evaluation (Recall, Precision, Accuracy, F1-Score)?



Model Building and Refinement

Work Package 3 | February 2021 – April 2022

Objective

Considering the results from Work Package 2, the objective in this Work Package is to build sustainable models that incorporate state-of-the-art NLP methods for Clinical Trials term coding

Main Activities

Medical History model development

- Based on selected approaches, build first MVP model and evaluate result
- Using hypotheses and evaluation criteria, refine model to improve accuracy
- Implement MedDRA version update

Adverse Events model development

- Based on selected approaches, build first MVP model and evaluate result
- Using hypotheses and evaluation criteria, refine model to improve accuracy
- Implement MedDRA version update

Automatic Integration of MedDRA Updates

- Develop approach for automatic integration of biyearly MedDRA Updates to reduce manual effort
- Test and validate results with Medical Coding Team

Involved Stakeholders

- Data Science Team(s)
- Project Manager (as required)
- Business Stakeholder (as required)

Outcomes

- Models for different Clinical Trials terms (MH&AE) that can be implemented in the Medical Coding processes
- Solution that enables automatic integration of MedDRA updates

Deep Dive: WP3 - User Requirements

Based on the Scope and Watson User Requirements, we derived requirements for the new system

User Requirements				
Category	URS No. (IBM Watson)	Requirement	Description	Status
Scope of Coding	SC-018	Propose Study Adverse Event terms	The new system must be able to propose coding of terms from the Adverse Event category for clinical studies to Low Level Terms in the MedDRA dictionary.	Req 1.0
Scope of Coding	SC-019	Propose Study Device Event terms	The new system must be able to propose coding of terms from the Clinical Event / Device Event category for clinical studies to Low Level Terms in the MedDRA dictionary.	Req 1.0
Scope of Coding	SC-020	Propose Study Medical History terms	The new system must be able to propose coding of terms from the Medical History category for clinical studies to Low Level Terms in the MedDRA dictionary.	Req 1.0
Scope of Coding	SC-024	Support terms in English	The new system must support terms submitted for coding in English.	Req 1.0
Knowledge Sources	KS-001.1	Coding dictionary	The new system must use MedDRA coding dictionary as a base knowledge source	Req 1.0
Knowledge Sources	KS-001.2	Bayer synonyms	The new system must use Bayer MedDRA synonyms as a base knowledge source	Req 1.0
Knowledge Sources	KS-001.3	Blocked terms	The new system must use Bayer MedDRA blocked terms as a base knowledge source	Req 1.0
Knowledge Sources	KS-001.14	Spell-checker	The new system must use a means of detecting spelling errors in terms submitted for coding as a base knowledge source	Req 1.0
Knowledge Sources	KS-001.15	Autocoded terms (MedDRA)	The new system must use a dataset of existing autoencoded terms for MedDRA as a base knowledge source	Req 1.0
Knowledge Sources	KS-001.19	Coded terms (MedDRA)	The new system must use a dataset of existing terms which once created omissions in MedDRA but now autoencode as a base knowledge source	Req 1.0
Proposing Solutions	PS-001	Acquire work	The interface must acquire batches of omissions from MPC.	Req 1.0
Proposing Solutions	PS-001.1	Prioritisation of work	MPC must apply the same prioritisation of work for the new system interface as it does for a human Proposer.	Req 1.0
Proposing Solutions	PS-002	Submit to Watson	The interface must submit records in an acquired batch to the new system for coding.	Req 1.0
Proposing Solutions	PS-003	Coding terms passed	The interface must pass to the new system the as-reported term and the alternate term.	Req 1.0
Proposing Solutions	PS-004.1	Coding terms passed	The interface must pass to the new system the Term Type	Req 1.0
Proposing Solutions	PS-004.2	Coding terms passed	The interface must pass to the new system the dictionary the new system must use to code the term	Req 1.0
Proposing Solutions	PS-004.3	Coding terms passed	The interface must pass to the new system the age of the subject (if available)	Req 1.0
Proposing Solutions	PS-004.4	Coding terms passed	The interface must pass to the new system the gender of the subject (if available)	Req 1.0
Proposing Solutions	PS-005	Terms to be evaluated	The new system must evaluate both the as- reported term and the alternate term when choosing a coding solution.	Req 1.0
Proposing Solutions	PS-006	Use of context items	The new system must evaluate the context items when choosing a coding solution.	Req 1.0
			For each record submitted by the interface, the new system must return its best five	

Discarded Requirements

The overall requirements of IBM Watson are still up to date and can be adopted for the new system. However the difference in scope of the new system with respect to PV and drug data results in an adjusted URS document. All requirements that remain for the new system have been aligned to the new scope and validated with business stakeholders.

Additional Requirements

As a result from various stakeholder interviews, an additional requirement has been added. The new system is therefore required to not simply make suggestions, as bad as they may be, but to also suggest to initiate a query to ask for additional information. In the long run, it should also be possible to initiate queries automatically once query formulations have been standardized.

The User Requirements build the foundation for the development phase with regard to features and functionalities. They need to be communicated to and aligned on with all team members and stakeholders.

Deep Dive: WP3 - Model Building and Refinement

There are different promising approaches that need to be evaluated carefully

Rule based

Process omissions created by AutoCoder to correct spelling errors and either feed them back to AutoCoder or process them separately.

Hypotheses

- AutoCoder cannot encode many terms due to spelling errors and abbreviations
- There are underlying and systematic errors within the omission data that can efficiently be addressed with additional rules

Methods

Text Mining
Fuzzy string matching
Utilizing Synonyms and Hyponyms
Pattern Recognition

Opportunities	Risks	Effort
Quick improvement of accuracy scores, Explainable and thus easier to adjust/finetune	Very static, might be incompatible with future MedDRA updates, too many rules required	Low, requires through understanding of underlying data to derive stable rules, fast initial implementation

Embedding Models

Use unsupervised machine learning to classify omissions utilizing embeddings and state-of-the-art language models and/or word distance measures.

Hypotheses

- There is too little context in input data to reasonably utilize embeddings
- Older models such as GloVe work might suffice

Methods

Embedding models (e.g., GloVe, BERT)
Deep Learning
Dimension Reduction
Distance measures (e.g., Word Mover's Distance, cosine similarity)

Opportunities	Risks	Effort
Might prove more stable with respect to MedDRA Updates, Embeddings can be trained task-specific	Too little context can lead to weak performance	Medium, needs advanced knowledge of Semantics & Deep Learning; fine-tuning of embedding models needed

Hierarchical Approach

Utilizing the MedDRA dictionary hierarchy to conduct a step-wise classification, e.g. using embedding models

Hypotheses

- The information provided in the MedDRA hierarchy is valuable and can be utilized
- Step-wise iterations improve classification accuracy due to less classes and dependencies

Methods

Utilizing hierarchy and ontology
Machine Learning
Bayesian tree classification

Opportunities	Risks	Effort
Available information in hierarchy can prove being valuable for classification	Very vulnerable to MedDRA hierarchy changes which can lead to high recalibration efforts	High, deep domain knowledge required, many iterations and models likely needed

Hybrid Models

Exemplary

Combine rule-based approaches with state-of-the art models to uplift accuracy and performance.

Hypotheses

- Combining different approaches iteratively will lead to better accuracy
- Adjusting weights between rules and semantics leads to similar performance as IBM Watson

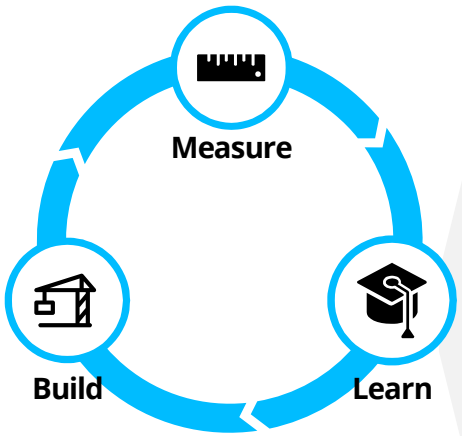
Methods

Rule based string matching
Embedding models
Distance measures
Ensemble learning

Opportunities	Risks	Effort
Combination of different approaches likely to increase overall accuracy and performance	Complexity increases while issues of all approaches need to be addressed together	High, due to iterative adjustments and complexity which requires time and both, broad and deep expertise

Deep Dive: WP3 - Model Building and Refinement

To refine the model in terms of accuracy, applying a iterative – build, measure, learn approach is key



Applying a iterative – build, measure, learn approach

Questions regarding NLP models & architecture:

Exemplary

- **Methods:** Which methods can be combined or stacked to improve performance?
- **Subclasses**: Which MedDRA classes perform best for which term groups (AE vs. MH)?
- **Performance**: Which model architectures perform best and why?
- **Architecture implications**: What kind of services is necessary for deploying the model (e.g. Docker, SageMaker, etc.) ?
- **Interface:** Which changes need to be implemented to the MPC interface – if any?
- ....

Feedback Loops with technical & NLP SMEs

Questions regarding User Requirements:

Exemplary

- **Features**: Which features from the User Requirements are to be prioritized and why?
- **Error patterns:** What kind of terms seem to be categorized wrong most often ?
- **Further data collection**: What further data points can be collected in order to offer value-added services?
- ....

Feedback Loops with Business SMEs

Architecture and Interfaces

Work Package 4 | April – October 2021

Objective

The objective of this Work Package is design and build a stable cloud environment for the new coding system including necessary interfaces.

Main Activities

Architecture & Infrastructure

- Review current IT architecture and assess Technical Requirements
- Draft future architecture of system, including interface to MPC & data pipeline
- Develop cloud infrastructure & maintenance processes

MPC Interface Development

- Based on architecture draft, develop, test and validate interface between MPC and new system
- Document required changes according to GxP validation process

Model Integration

- Build model-pipelines (including re-training loops) and integrate model in cloud environment
- Test and validate model pipeline
- Provide documentation for architecture, model and pipelines

Involved Stakeholders

- Lead Architect
- ML/Cloud Engineer
- IT System Owner
- Project Manager
- Validation SME
- Business Stakeholder (as required)

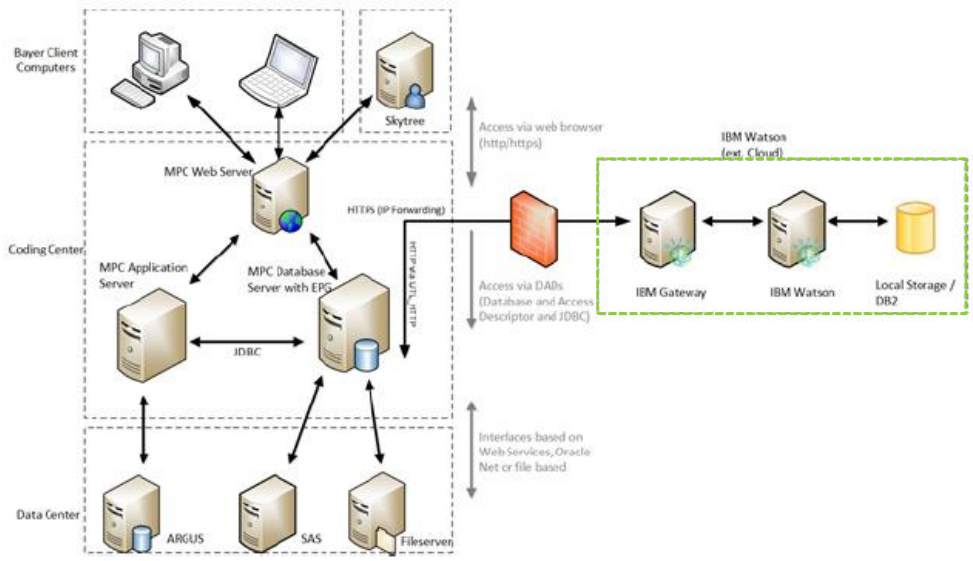
Outcomes

- Cloud based infrastructure that is in line with User Requirements and internal IT requirements (e.g. GxP)
- Documentation of different components such as pipelines, architecture, etc.
- Validated interface between MatchPoint Coder and new system

Deep Dive: WP4 - Architecture and Interfaces

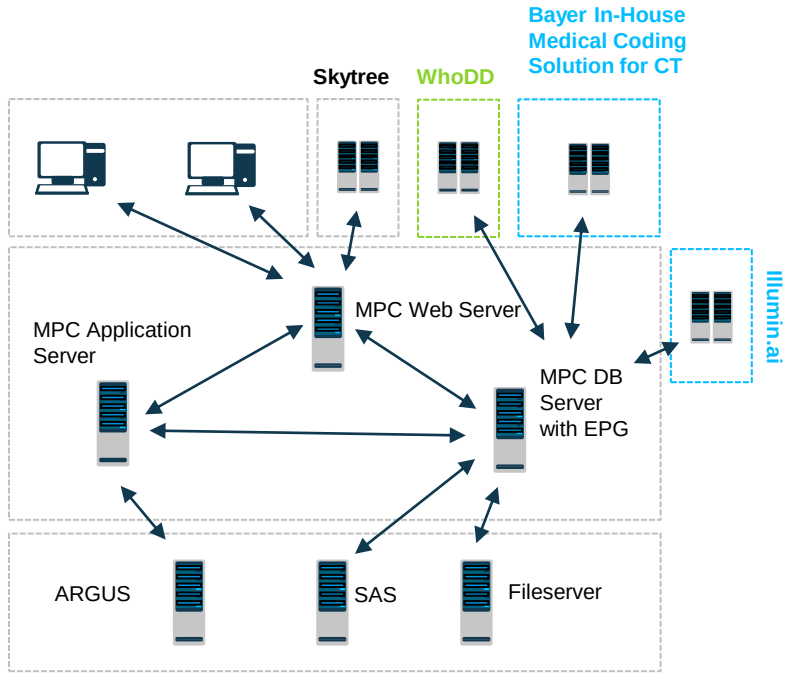
Due to the planned replacement of IBM Watson, changes to the current architecture will be necessary

Today



- Today, coding of CT, PV and drug data is conducted via IBM Watson interface only
- For Japanese terms, Skytree is being used

2022



- New Bayer In-House Medical Coding is proposing solutions for CT terms only
- PV data is coded via Illumin.ai system
- WhoDD API is used for drug data coding
- For Japanese terms, Skytree is still in place

External System

New Bayer System

Testing and Deployment

Work Package 5 | April – November 2021

Objective

Ensure MatchPoint Coder system is compliant after integration of the new system. This includes all relevant testing and validation efforts according to compliance standards as well as a successful deployment and Go-Live.

Main Activities

Integration and Validation Testing

- Develop test plans for integration and validation testing based on internal best-practices and requirements
- Conduct integration and validation testing
- Evaluate and document results according to GxP standards

User Acceptance Testing

- Develop test plan and prepare User Acceptance Testing considering stakeholder expectations and availabilities
- Conduct User Acceptance Test
- Evaluate and document results

Deployment and Go-Live

- Develop cutover and roll back plans for Go-Live
- Request and gather internal reviews and approvals
- Prepare and manage Go-Live

Involved Stakeholders

- Data Science Team(s)
- ML/Cloud Engineer
- IT System Owner
- Project Manager
- Validation SME
- Business Stakeholder (as required)

Outcomes

- Successfully performed UAT as well as integration and validation testing
- Successful Deployment and Go-Live of new system
- Complete set of validation documents

Hypercare and Documentation

Work Package 6 | December 2021 – May 2022

Objective

Main focus of Work Package 6 are downstream tasks such as successfully disconnecting IBM Watson from the coding environment, monitoring the new system and close the project. The objective is to have a system that runs stable and a well-documented project closure.

Main Activities

Decommissioning of IBM Watson

- Shutdown IBM Watson interface

Hypercare of new system

- Monitor new system performance and operations
- Manage issues and bug fixes as required

Project Documentation

- Document Lessons Learned
- Initiate official Project Closure

Involved Stakeholders

- Data Science Team(s)
- Project Manager
- IT System Owner
- ML/Cloud Engineer
- Validation SME
- Business Stakeholder (as required)

Outcomes

- All infrastructure components regarding IBM Watson are successfully decommissioned
- The new system runs stable
- Project documentation is compliant with all internal requirements (such as GxP)

Team Overview & Required Profiles

Team Overview

Project Management	**Project Lead** Sebastian Nedo	**Product Owner** Martina Viell	**NLP Lead** Dr. Marion Schwärzler
Data Science	**Lead Data Scientist** Angelo	**NLP Researcher MH** Enrico	**NLP Researcher AE** tbd
Engineering	**Lead Architect** tbd	**Cloud/ML Engineer** tbd	
SMEs	**Business SME** Hanna Viol	**Validation SME** tbd	**NLP SME** Prof. Dr. Alan Akbik
	Process Manager Tanja Bellaire (tbc)	**System Owner** Amit Pathak	**MPC SME** Christoph Brieden

Bayer External

*Names to be confirmed

Team Overview: Required Profiles

Project Management

Project Management

Data Science

Engineering

SMEs

	 **Project Lead** Sebastian Nedo	 **Product Owner** Martina Viell	 **NLP Lead** Dr. Marion Schwärzler
Main responsibilities	• Responsible for delivering the project in time and budget considering the scope, • Lead of WPs, 1, 5 & 6	• Take part in steering committee meetings • Ensure Business SME availability • Provide feedback to DS streams in model development	• Take part in steering committee meetings • Accountable for Data Science stream results and deliveries
Experience Level	• > 5 years experience in managing projects and implementing software solutions, ideally at Bayer	• n/a	• n/a
Required Skills	• Extensive knowledge of Project Management approaches, Bayer validation processes and technical aspects	• n/a	• n/a
Estimated Efforts	• 330 PD	• 39 PD	• 36 PD

Team Overview: Required Profiles

Data Science

Project Management

Data Science

Engineering

SMEs

	 **Lead Data Scientist**	 **NLP Researcher**	 **NLP Researcher**
Main responsibilities	• Lead of WPs 2 & 3 • Manage model development teams • Develop NLP models • Align model development with project goals and enable knowledge transfer	• Team Member of CT-MH model development stream • Research and evaluate NLP methods • Develop, test and validate model results with respect to requirements	• Team Member of CT-AE model development stream • Research and evaluate NLP methods • Develop, test and validate model results with respect to requirements
Experience Level	• > 5 years experience in developing and deploying NLP models • Experience in managing Machine Learning Teams	• > 3 years experience in developing NLP models	• > 3 years experience in developing NLP models
Required Skills	• Deep understanding of NLP and ML methods • Strong programming skills (Python) • Strong organizational, communication and leadership skills	• Deep understanding of NLP and ML methods • Strong programming skills (Python) and experience with ML frameworks such as Tensorflow, Keras or Pytorch	• Deep understanding of NLP and ML methods • Strong programming skills (Python) and experience with ML frameworks such as Tensorflow, Keras or Pytorch
Estimated Efforts	• 203 PD	• 252 PD	• 251 PD

Team Overview: Required Profiles

Engineering

Project Management

Data Science

Engineering

SMEs

	 **Lead Architect**	 **ML Engineer**
Main responsibilities	• Lead of WP 4 • Analyze technology environment and infrastructure specifics • Identify dependencies & draft architecture • Support testing and validation activities	• Team Member of WPs 4 & 5 • Set up ML infrastructure and deploy NLP models for Medical Coding • Design data pipelines and interfaces
Experience Level	• > 8 years designing and building IT infrastructures and cloud architectures • > 5 years experience in leading and managing Engineering teams	• >3 years of experience in building ML infrastructure and deployment ready applications related to natural language processing and machine learning
Required Skills	• Deep experience in cloud and software architecture design as well as DevOps • Excellence communication skills • Collaborative work approach and analytical skills	• Deep knowledge of Cloud infrastructure services (AWS, etc.) • Knowledge of DevOps concepts • Experience with ML/DL frameworks such as Tensorflow, Keras or Pytorch,
Estimated Efforts	• 179 PD	• 161 PD

Team Overview: Required Profiles

Subject Matter Experts

Project Management

Data Science

Engineering

SMEs

	Main responsibilities	Estimated Efforts
 **Business SME**	• Provide feedback during model development regarding performance and error patterns • Support DS with business knowledge	• 74 PD
 **System Owner**	• Responsible for MPC system, including availability, installation and maintenance • Supports validation and deployment of new system with respect to MPC	• 15 PD
 **Process Manager**	• Ensuring that the system and its operation is fit for intended use and in compliance with all applicable regulations	• 15 PD
 **MPC SME**	• Support different teams by providing data, feedback and knowledge about MPC interface as well as during validation and deployment	• 122 PD
 **NLP SME**	• Support Data Science Streams in Sprint planning and reviews through providing feedback and suggestions with regard to model development	• 19 PD
 **Validation SME**	• Ensures compliance with applicable requirements and policies • Reviews and approves system deliverables and releases system	• 131 PD

Business Case

Business Case

The cost and effort estimates result in a positive Business Case, further supported by qualitative benefits

Quantitative

Cost	2021	2022	2023	2024	2025	2026	Total
One time cost	1.321.700 €	263.700 €					**1.585.400 €**
Run cost		60.000 €	60.000 €	60.000 €	60.000 €	60.000 €	**300.000 €**
Total*	**1.321.700 €**	**323.700 €**	**60.000 €**	**60.000 €**	**60.000 €**	**60.000 €**	**1.885.400 €**

*Not included in estimates are one-time and run cost for cloud resources as they are controlled centrally. They are estimated with ~40.000€ per year.

Benefits	2021	2022	2023	2024	2025	2026	Total
Additional revenue	- €	- €	- €	- €	- €	- €	**- €**
Savings	- €	532.995 €	532.995 €	532.995 €	532.995 €	532.995 €	**2.664.975 €**
Cost Avoidance		79.949 €	79.949 €	79.949 €	79.949 €	79.949 €	**399.746 €**
Total							**3.064.721 €**

Pay-Back Period (in years) **3,31**

Net Present Value **489.207 €**

Qualitative

- Support Digital Roadmap through expanding Bayer's internal AI skills and expertise
- Ensure available Business IP stays within Bayer and further develop new IP, i.e. AI algorithms and/or patents
- Utilize most recent advances in NLP domain to increase accuracy for medical coding
- Enable future use cases such as medical coding of Real World Data
- Reduce dependency from external vendors to allow for further development and improvement

Next Steps

Next Steps

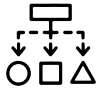
With the proposed project start being only 6 weeks away, internal initialization has to start now

Request and secure budget for 2021 and 2022 to avoid delays in project delivery

Determine available employee resources and **identify** needs for external support

Involve relevant stakeholders and **close** information gaps (e.g. regarding cloud infrastructure)

Gather relevant AutoCoder and omissions data and identify further data sources



Prepare project kick-off and set up project management

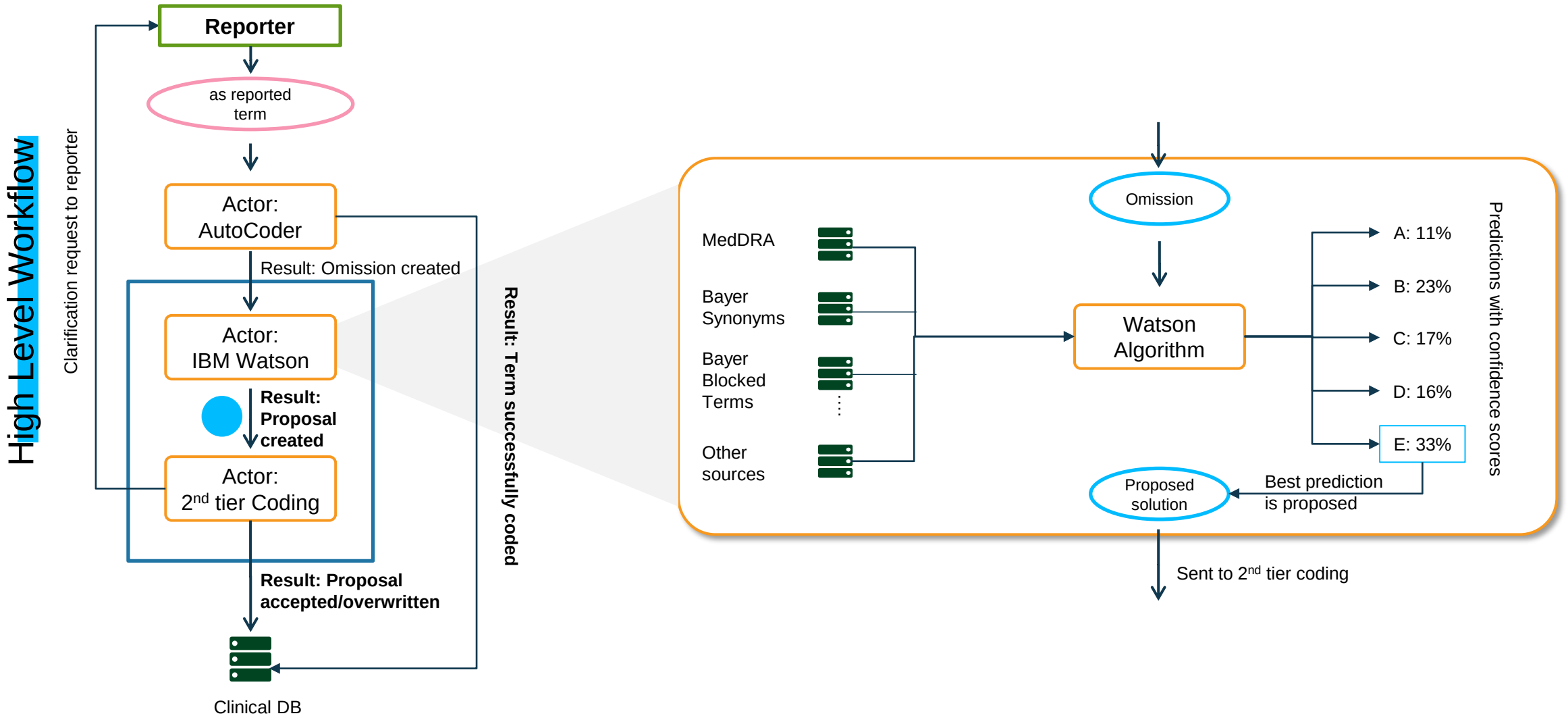
Validate User Requirements and Effort Estimates, adjust when necessary

Initiate further initiatives required for a full replacement of IBM Watson (e.g. UMC Drug Coding PoC)

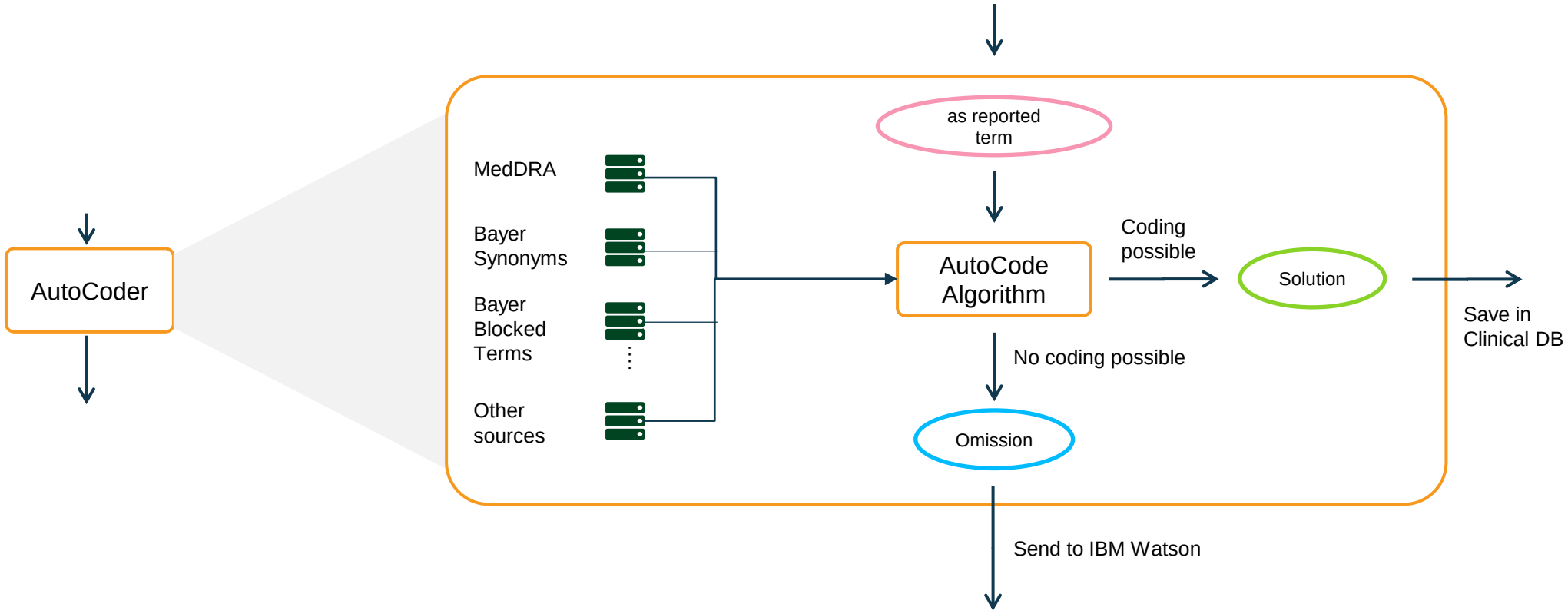
Start negotiations with IBM regarding a 1-year extension to have a fallback option and sharpen Business Case

Appendix

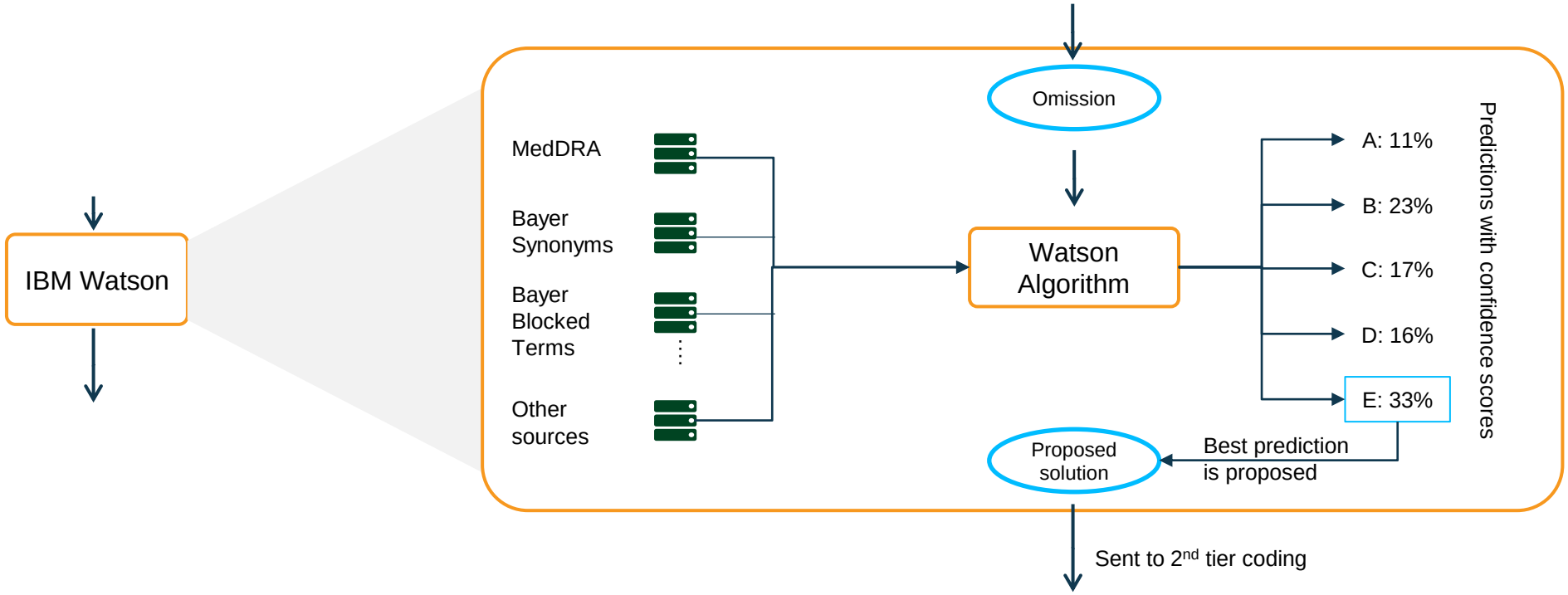
High Level Workflow



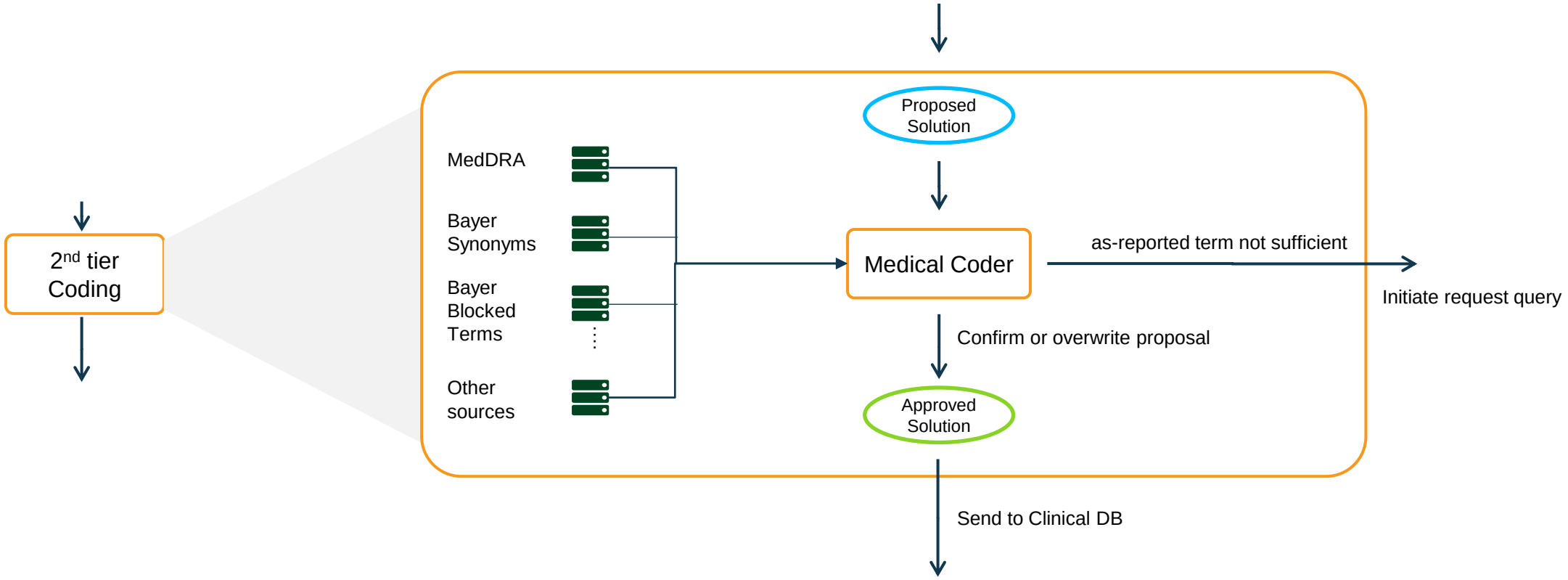
High Level Workflos – Zoom-in: AutoCoder



High Level Workflos – Zoom-in: IBM Watson



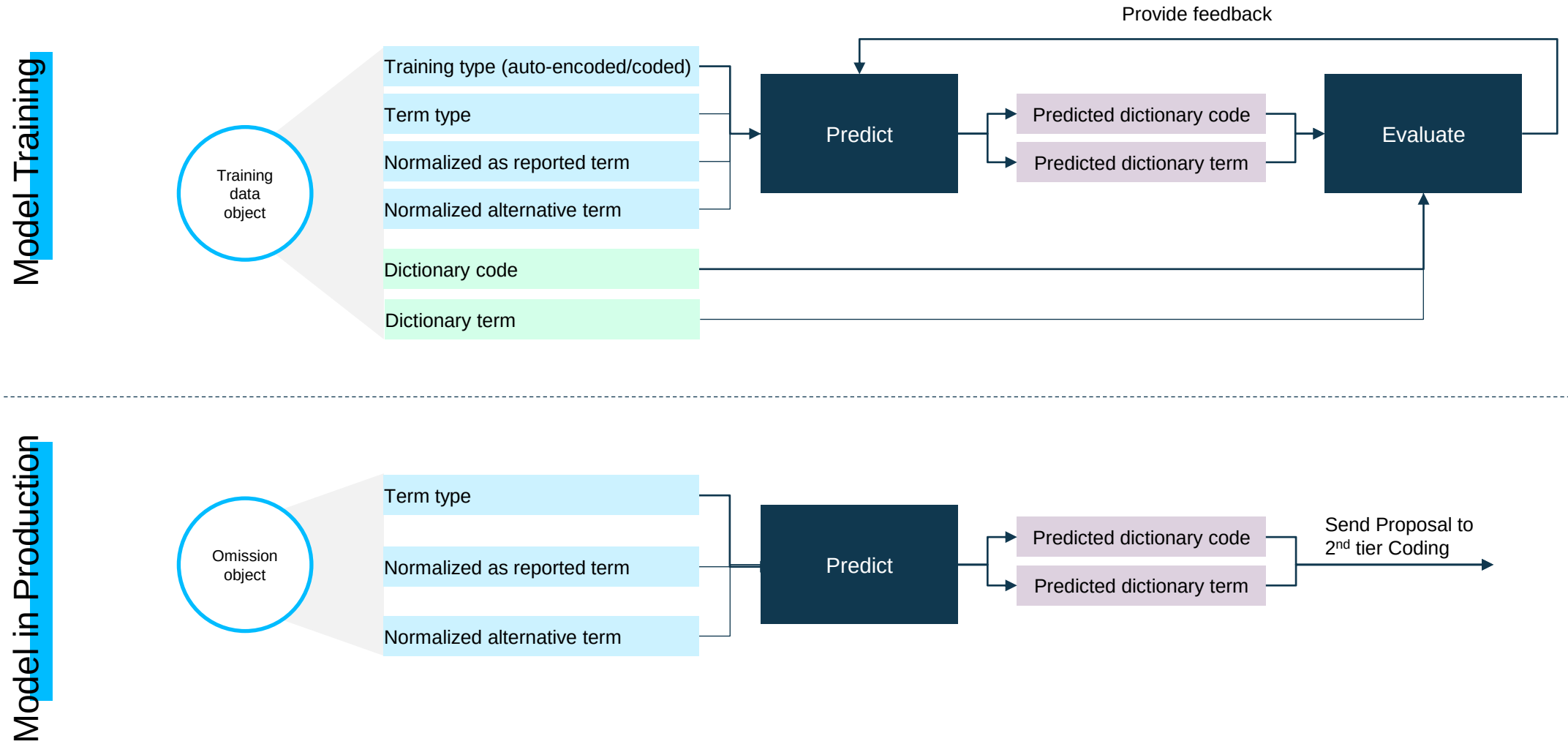
High Level Workflos – Zoom-in: 2nd tier Coding



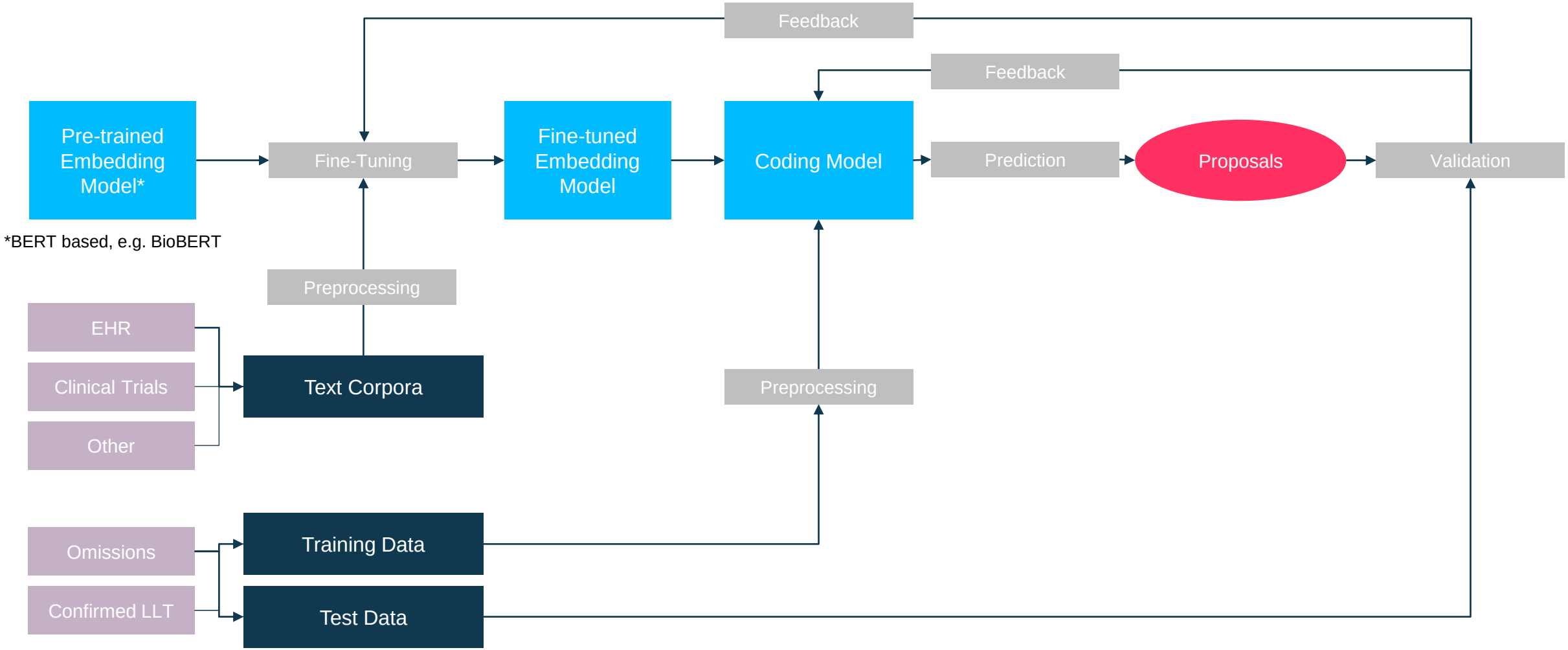
Current data model

```
<DATA_RECORD>
  <STUDY_NAME>17613</STUDY_NAME>
  <TERM_TYPE_NAME>CE</TERM_TYPE_NAME>
  <RECORD_ID>20170718145126284109050045022</RECORD_ID>
  <REP_TERM>LEFT EYE SURGERY: CATARACT SURGERY</REP_TERM>
  <ALT_TERM></ALT_TERM>
  <LLT_CODE>10063797</LLT_CODE>
  <LLT_NAME>Cataract operation</LLT_NAME>
  <PT_CODE>10063797</PT_CODE>
  <PT_NAME>Cataract operation</PT_NAME>
  <AUTOCODE_DATE>2019/10/24 07:23:24</AUTOCODE_DATE>
</DATA_RECORD>
```

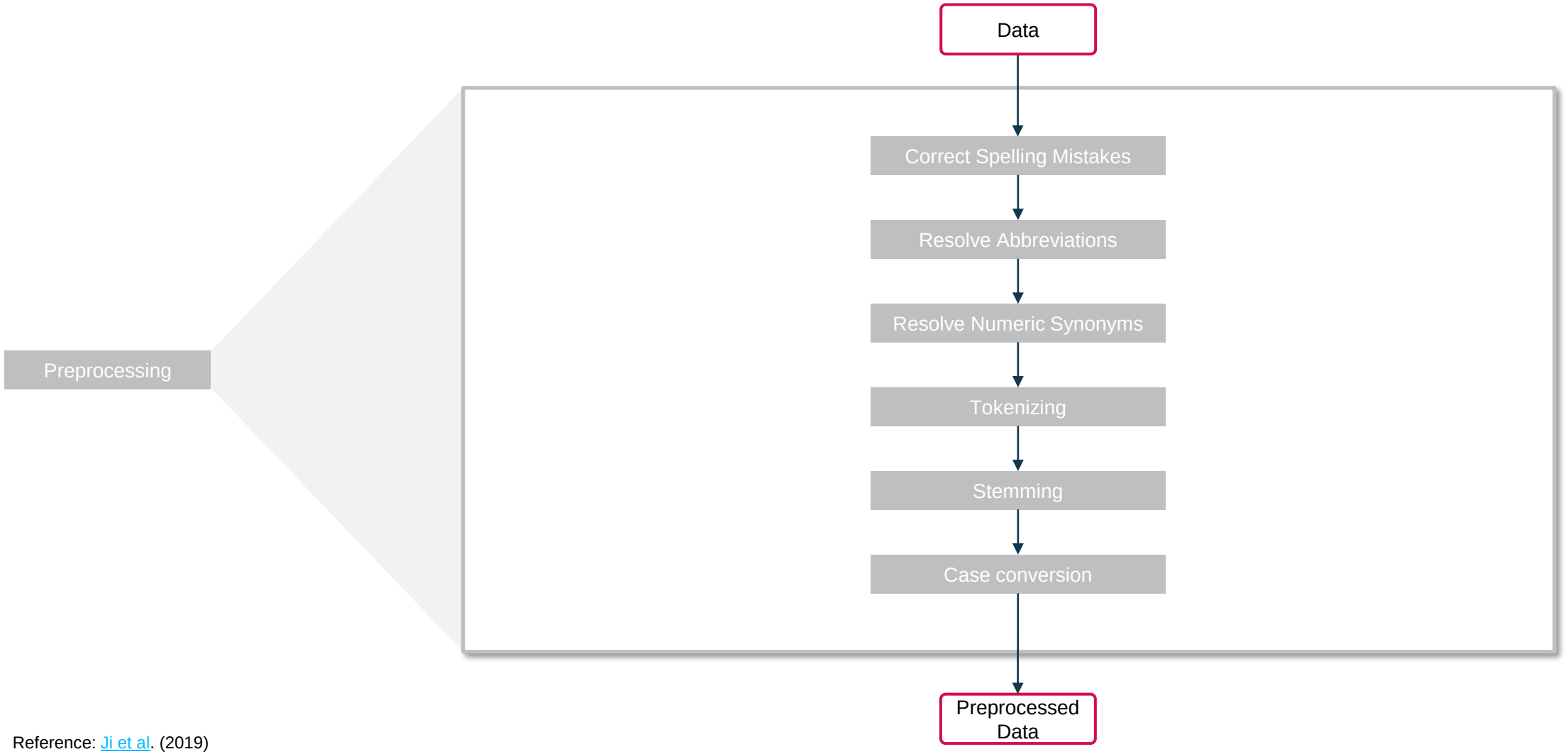
Current understanding of IBM Watson training & prediction



Possible Model Architecture for Embedding Models

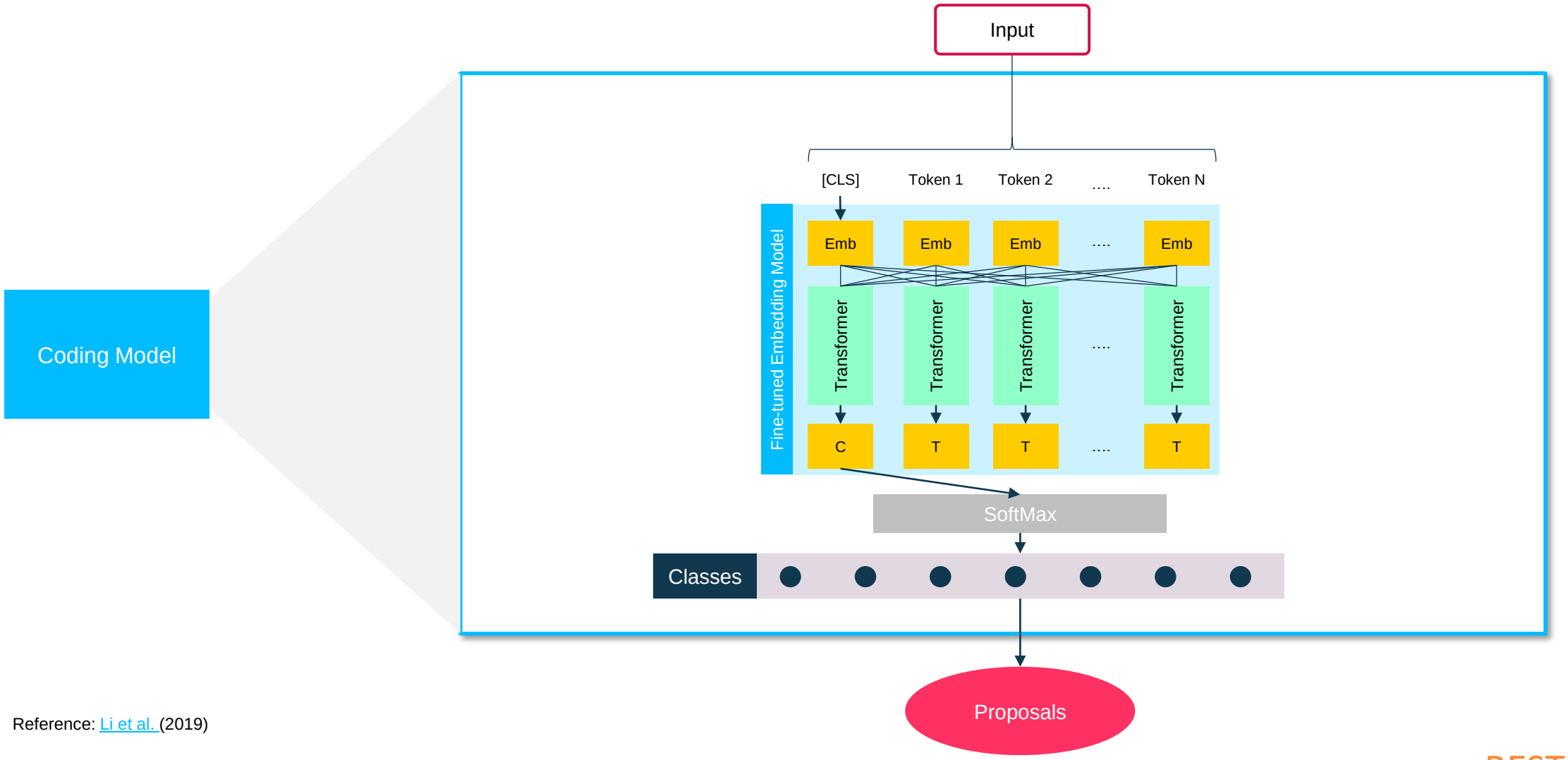


Possible Model Architecture – Zoom-In: Preprocessing



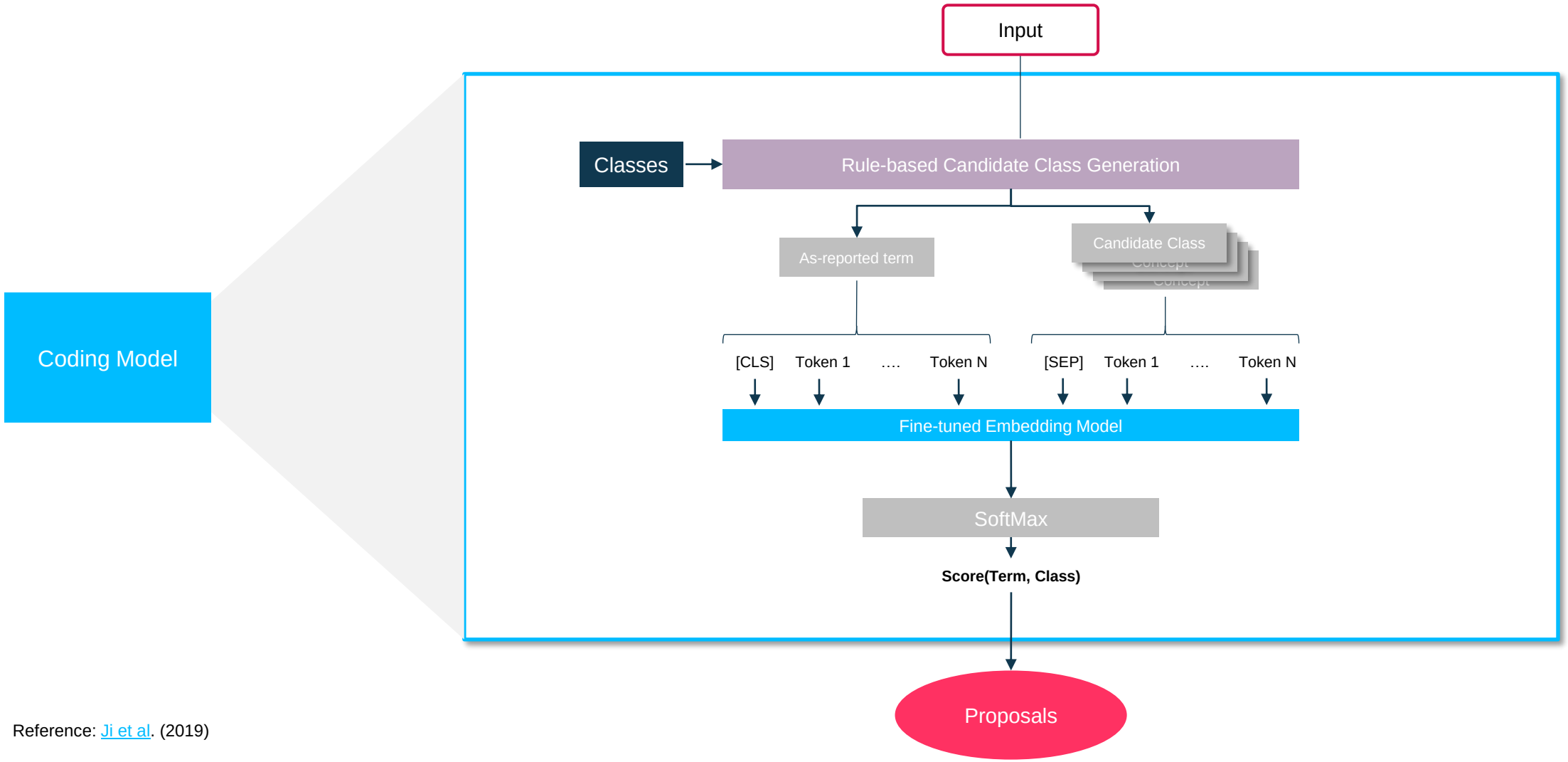
Reference: Ji et al. (2019)

Possible Model Architecture – Transformer Classification Model



Reference: Li et al. (2019)

Possible Model Architecture – Hybrid Classification Model

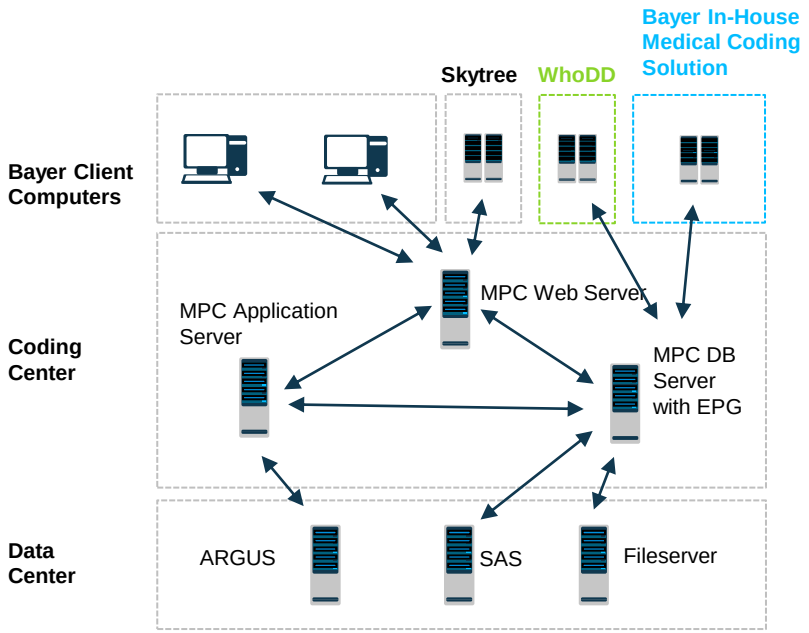


Reference: Ji et al. (2019)

Options: Future architecture

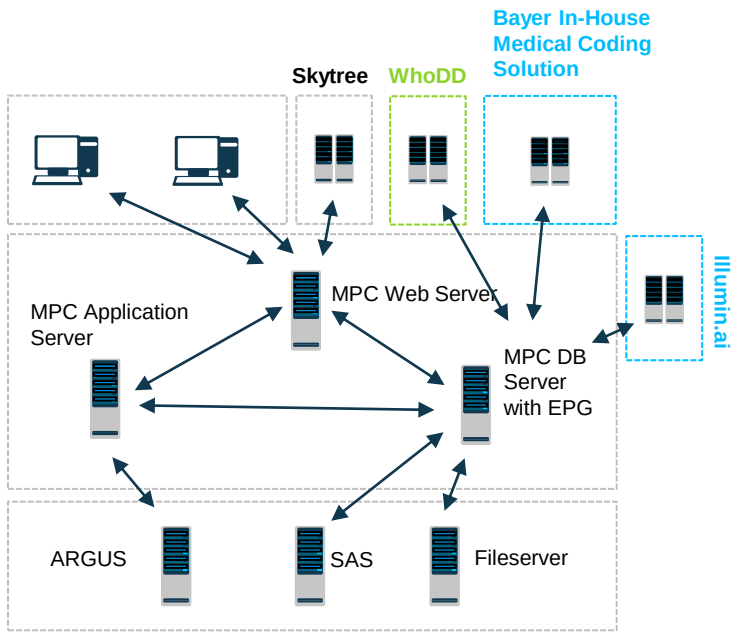
There are different options for the future coding architecture, depending on the scope

Option 1



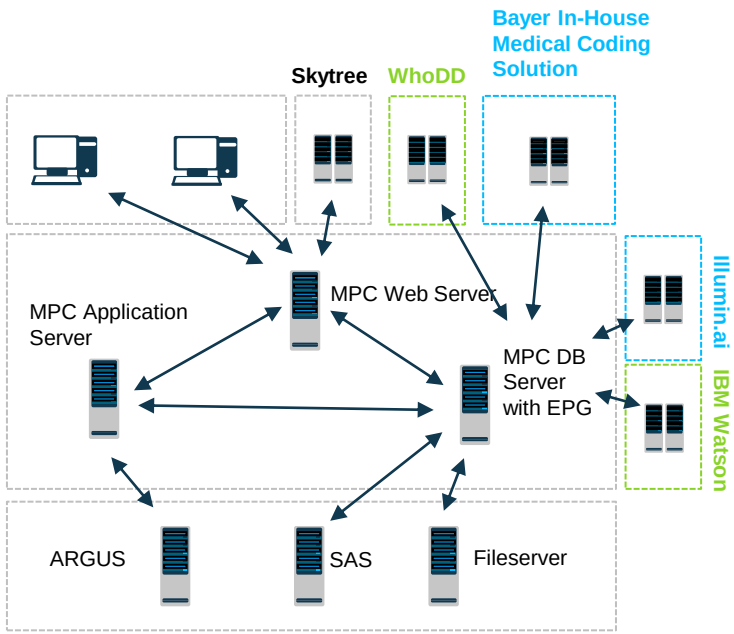
- New Bayer In-House Medical Coding is proposing solutions for both, CT and PV terms
- WhoDD API is used for drug data coding
- For Japanese terms, Skytree is still in place

Option 2



- New Bayer In-House Medical Coding is proposing solutions for CT terms only
- PV data is coded via Illumin.ai system
- WhoDD API is used for drug data coding
- For Japanese terms, Skytree is still in place

Option 3



- New Bayer In-House Medical Coding is proposing solutions for CT terms only
- PV data is coded via Illumin.ai system
- IBM Watson is still being used for extra coding loops
- WhoDD API is used for drug data coding
- For Japanese terms, Skytree is still in place

New external System

New Bayer Sytem