

Comissionamento e controle de qualidade de sistemas baseados em IA

Exemplos e Cuidados

Commissioning and QA of AI-based systems

Examples and Cautions



DukeHealth

Justus Adamson, PhD

Associate Professor

Duke Radiation Oncology



Conflict of Interest Statement

- Research funding from Radialogica LLC
- Ownership of Clearsight RT LLC (unrelated to this topic)

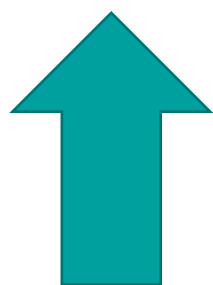


Inteligência Artificial (Artificial Intelligence or AI):

Uma abordagem computadorizada para **identificar relações matemáticas complexas** dentro de **dados observacionais**, e que pode **evoluir** à medida que novos dados são gerados.

A computerized approach to **identify complex mathematical relationships** within **observational data**, and that can **evolve** as new data are generated.

Porque Agora? (Why is this is AI's moment?)



Capacidade de Computação (Computing ability)

- Processadores mais rápidos
(Faster processors)
- Processamento paralelo
(Parallel / GPU processing)
- Cloud based processing
- Algoritmos melhorados
(Improved algorithms)



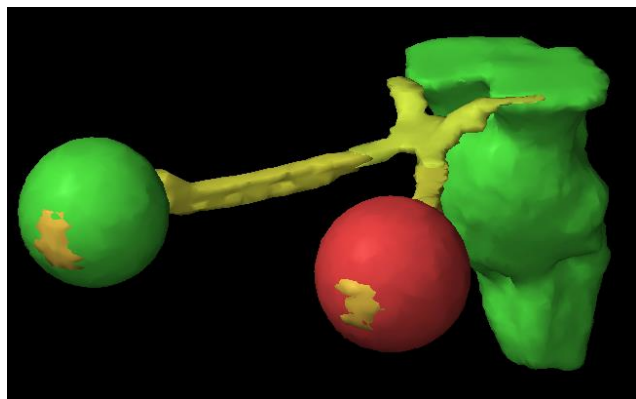
Disponibilidade de Dados (Data availability)

- Coleta de dados automatizada
(Automated data collection)
- Prontuário eletrônico (EMR)
- Computerized treatment planning
- Mais Imagens (Increased imaging)
- Resultado da terapia (Outcome data)
- Dados de Patologia (Pathology data)
- Dados Genéticos (Genetic data)

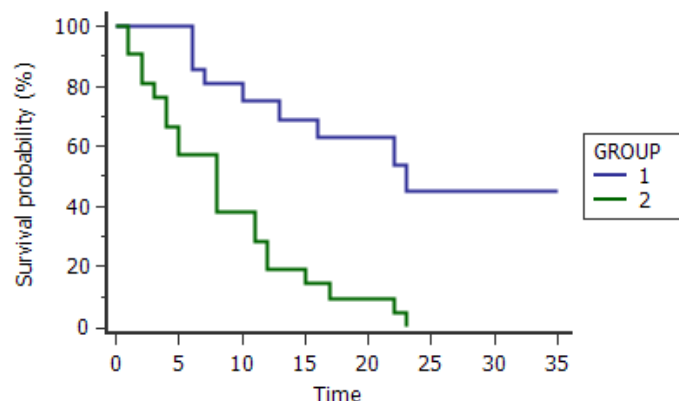


AI Applications in Radiation Oncology:

Segmentação automática (auto-contouring)



Seleção assistida (decision support)

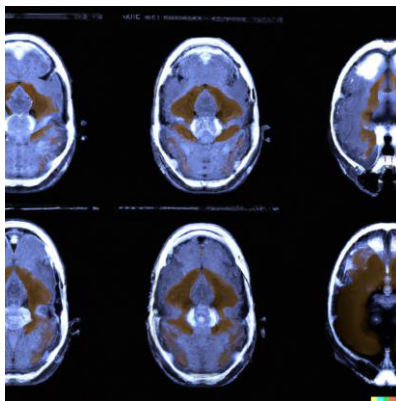


Previsão de (Prediction of):

- Plan specific dose tradeoffs
- Resultados derivados de dose e outros dados Outcomes from dose & orthogonal data (pathology, etc.)
- Resultados derivados de imagens Outcomes from imaging (Radiomics)

Manipulação de imagens (image manipulation)

- Synthetic CT / MRI
- image quality improvements



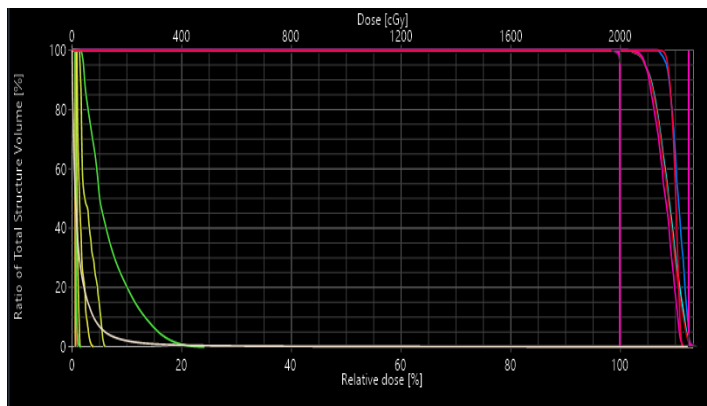
Qualidade (QA):

Previsão de (Prediction of):

- IMRT-QA passing rates
- Image guidance errors
- Linac performance (preventative maintenance)



Planejamento automático (auto-planning)





AI Applications in Radiation Oncology:

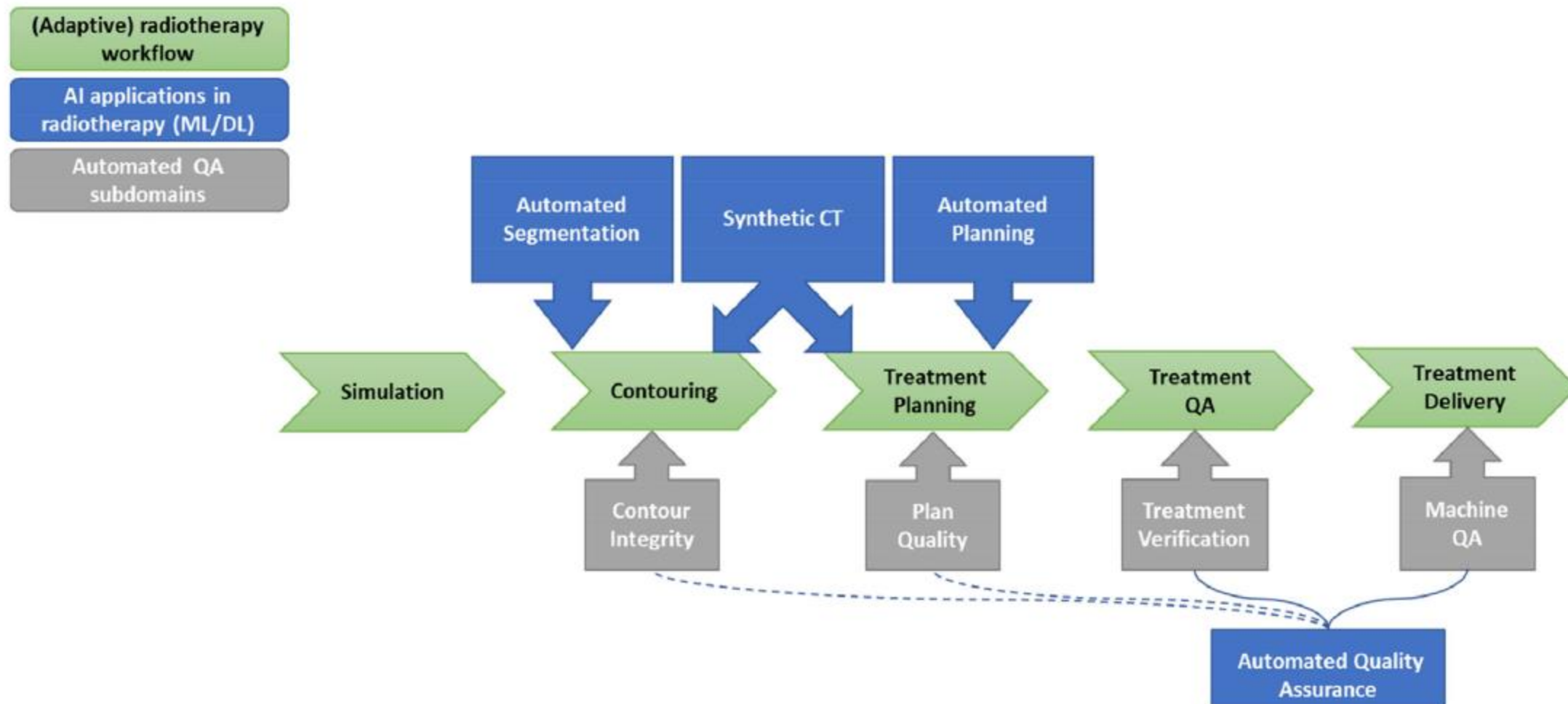


Fig. 1. Current use of selected AI applications in the radiotherapy workflow discussed in this paper. Other workflow steps that could benefit from AI such as image reconstruction, registration, etc. that are not discussed in this paper are not shown in this figure.

Por que os sistemas de
IA precisam de controle
de qualidade?

Why do AI systems need
QA?

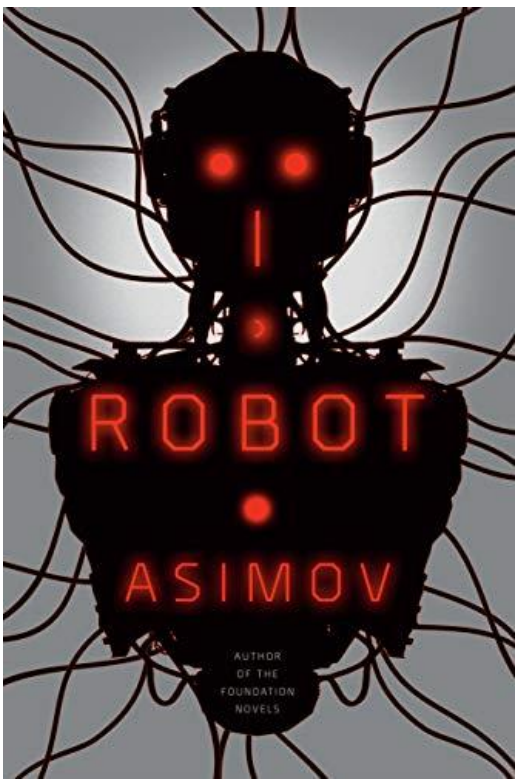


DukeHealth



História: controle de qualidade para IA (Background)

1950



Isaac Asimov publica “Eu, Robô” que prevê “3 leis da robótica:

Asimov publishes “I Robot” which envisions “3 laws of robotics:

1. Robôs não devem prejudicar humanos.

Robots shall not harm humans.

2. Robos devem obedecer humanos.

Robots must obey humans

3. Robôs devem evitar ações que possam causar danos a si mesmos

Robots must avoid actions that could cause it to come to harm

Como o conceito de inteligência artificial de Asimov se sustenta hoje?

How does Asimov’s concept of artificial intelligence hold up today?



Por que os sistemas de IA precisam de controle de qualidade?

Why do AI systems need QA?

2016

 Interesting Engineering

Microsoft's Ai Bot Turned Racist After Hours on the Internet

Developed by Microsoft's Technology and Research team, Tay, with the twitter handle @Tayandyou, is targeted at U.S. 18-24 year olds which...

Mar 25, 2016



2021

 WIRED

It Began as an AI-Fueled Dungeon Game. It Got Much Darker

In December 2019, Utah startup Latitude launched a pioneering online game called AI Dungeon that demonstrated a new form of human-machine...

May 5, 2021



2022

 CNN

It didn't take long for Meta's new chatbot to say something offensive

... such as with Microsoft's "Tay" chatbot more than six years ago. The colorful responses from BlenderBot show the limitations of building...

6 hours ago



August 10th Last week!



Um exemplo ilustrativo: (An illustrative example)

DALL·E 2



DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.

INSTRUÇÕES PARA AI: (PROMPT)

“Uma fotografia de um guaxinim que é um astronauta.”

“A photograph of a raccoon who is an astronaut.”





DALL-E2 Ejemplos: (Examples)

INSTRUÇÕES PARA AI: (PROMPT)

"uma radiografia de um abdome"

"a radiograph of an abdomen"



INSTRUÇÕES PARA AI: (PROMPT)

"Um panda estudando uma radiografia em um hospital."

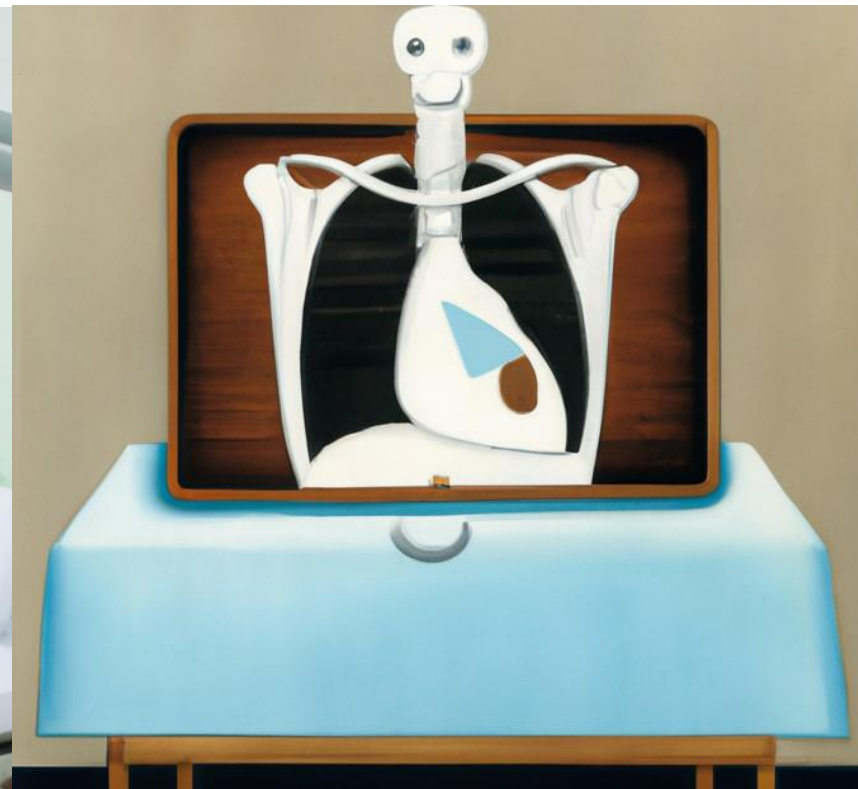
"A panda looking at a radiograph in a hospital."



INSTRUÇÕES PARA AI: (PROMPT)

"uma pintura de Tarsila do Amaral de uma radiografia de tórax"

"Tarsila do Amaral painting of a chest radiograph"





Exemplos de Erros: (Examples of failures)

INSTRUÇÕES PARA DALL-E2

AI: (PROMPT FOR AI)

"a clinical linear accelerator".

Not a great design for the patient support couch





Exemplos de Erros: (Examples of failures)

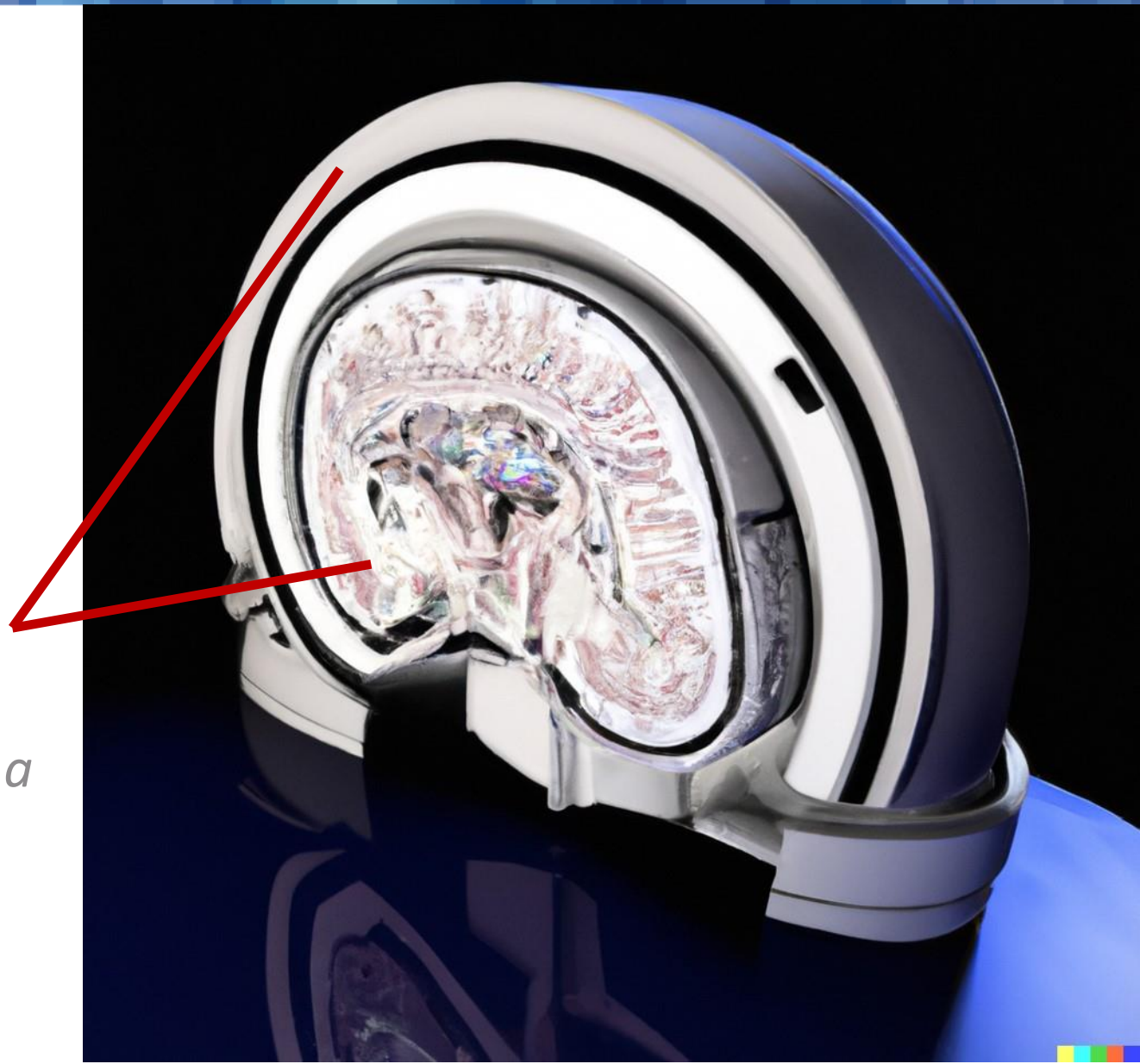
INSTRUÇÕES PARA DALL-E2

AI: (PROMPT FOR AI)

“a 3d rendering of a computed tomography image”

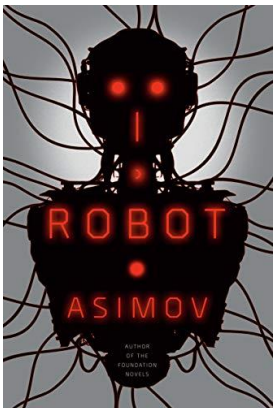
*O algoritmo não sabia se deveria desenhar um **scanner** de TC ou uma **imagem** de TC, então os combinou?*

*Algorithm couldn't decide whether to draw a **CT scanner** or a **CT image**, so it combined them?*





Lições Aprendidas: (Lessons learned)



1. Asimov estava errado (Asimov was wrong)

Asimov assumed the biggest challenge would be building the AI, while defining appropriate limits would be trivial. The opposite is true: now we can easily build AI, but defining appropriate limits & safety features is a current challenge.



2. Dados precisos são essenciais. (Good data is essential)

It's much easier to **curate training data upfront** than to police outcomes afterwards.



3. Maior sofisticação requer maior escrutínio.

(Greater sophistication requires greater scrutiny)

More sophisticated methods (deep learning) have outsized potential & broader application(s) but are **more difficult to know when they have exceeded their limitations.**

Aplicação em Radioterapia

Applying this to
Radiation Oncology



DukeHealth

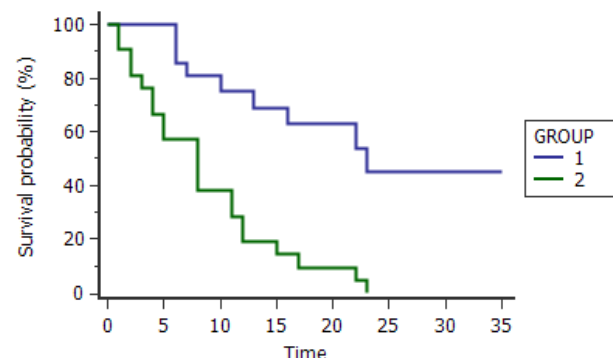
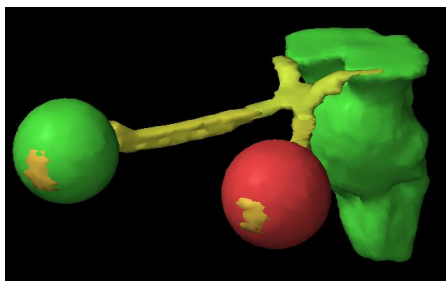


Por que os sistemas de IA precisam de controle de qualidade?

Why do RadOnc AI systems need QA?

Consequências Potenciais de Falhas: (Potential Consequences of Failures)

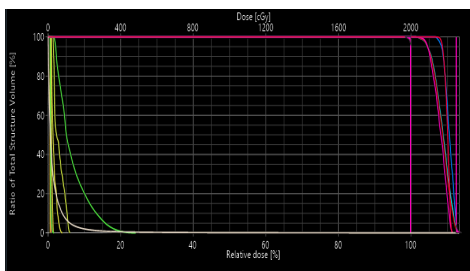
- Incorrect PTV / OAR definitions



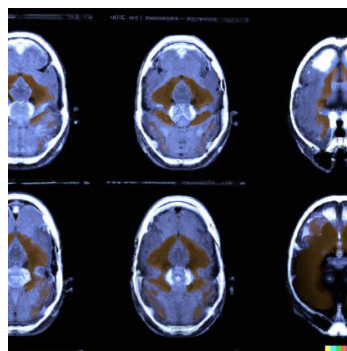
- Sub-optimal selection of of planning technique / treatment dose
- Incorrect outcome prediction

time cost / inefficiency due to need to “re-do” AI task(s) or need for increased effort to verify AI results

- Poor quality treatment plans



- Incorrect image details
 - Increased risk of mis-alignment
- Sub-optimal image quality



- False positives / negatives for predictions of IMRT QA, image guidance errors, & linac performance

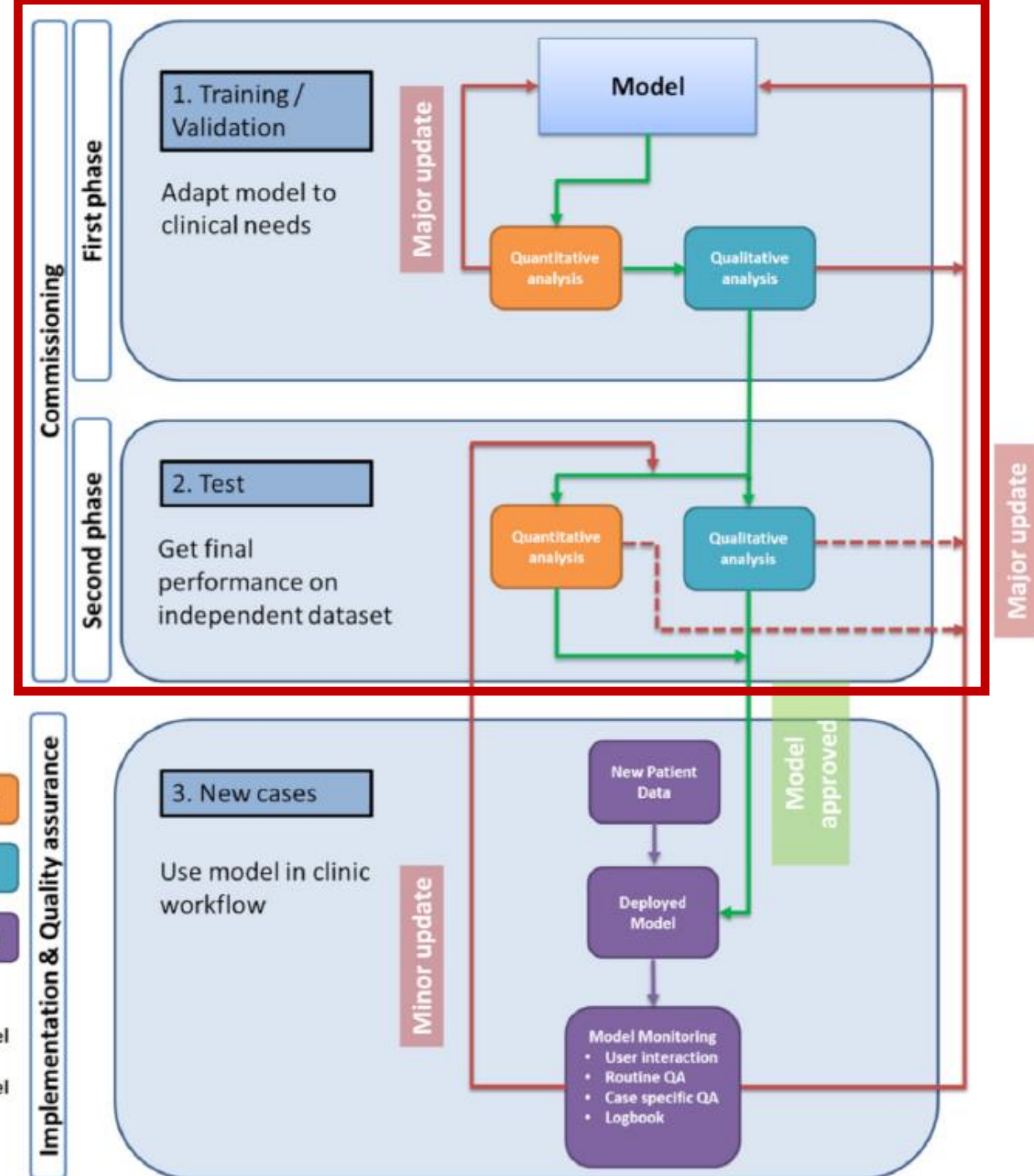


Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." Radiotherapy and Oncology 153 (2020): 55-66.

Commissioning (prior to clinical use)

- Train AI model (for in house models)
- Investigate accuracy & reproducibility
- Two phases:
 - Training/validation
 - (Independent) testing

Legend





Commissioning Guidelines: Training Step

- Ensure appropriateness & accuracy of training data (important!!)
 - Is the data accurate? up to date? relevant?
 - Should be a curated representation of the patient population and clinical practice under consideration, which can require triage to generate reliable subsets samples.
 - What are the limitations of the data?
 - Does it include important rare cases / failures?
 - Example: patients with rare anatomical variations for auto contouring (such as hypertrophic prostate median lobe)
 - Example: failure cases for IMRT QA prediction models
 - Does it include the full range of variables expected to be encountered during clinical use?
 - Is there sufficient data to train a robust model?
 - Should include enough data. *However*, well-known, not too large, high quality datasets are generally preferred over very large datasets of lower quality due to the evolving nature of clinical protocols and guidelines

Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." *Radiotherapy and Oncology* 153 (2020): 55-66.



Commissioning Guidelines: Validation Step

- Validation performed with a smaller set of data (validation dataset)
 - Often the data is randomly divided-> 80/20% split, etc.
- Should represent the data on which the model will be applied clinically
- Evaluate performance using a range of metrics (in addition to those used for training)
- Review performance with relevant clinical experts (physicians, physicists, radiation therapists, etc.)

Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." *Radiotherapy and Oncology* 153 (2020): 55-66.



Commissioning Guidelines: Commercial Systems

- In some cases, training / validation is still carried out in house (data trained on institutional data)
- More often, the model is trained by the company prior to delivery to the institution.
 - Validation step already completed / not performed by institution
- Institutional responsibility:
 - Verify proof of FDA / CE mark
 - Review vendor's validation report detailing the performance of the model
 - Important aspects: intended use and limits, description of training and validation set, used standards (if applicable), metrics and overall validation protocol

Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." *Radiotherapy and Oncology* 153 (2020): 55-66.



Commissioning Guidelines: Test Phase

- Goals of test phase:
 - Independent evaluation of the final performance of the model
 - Investigate the robustness of the model
 - Define for which (type of) patients the model can be applied
- Test phase dataset:
 - Should be Independent
 - Should represent the data for which the model will be used clinically
 - Should show similar variation as in the training data
 - Minimum dataset size:
 - no clear cut answer
 - may depend on variation in input data; min of 10 recommended as a good starting point
- Quantify model performance, but no changing of the model based on these results.
 - If model performs poorly, move back to training / validation step.

Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." *Radiotherapy and Oncology* 153 (2020): 55-66.



Commissioning Guidelines: Other Tasks

- Risk assessment (such as FMEA)
- Develop clinical workflow that ensures safe & clinically relevant use
 - Use risk assessment in developing clinical workflows
- Provide training to users for correct use and interpretation of model output
- End to end test
- Involve a interdisciplinary commissioning team for these tasks

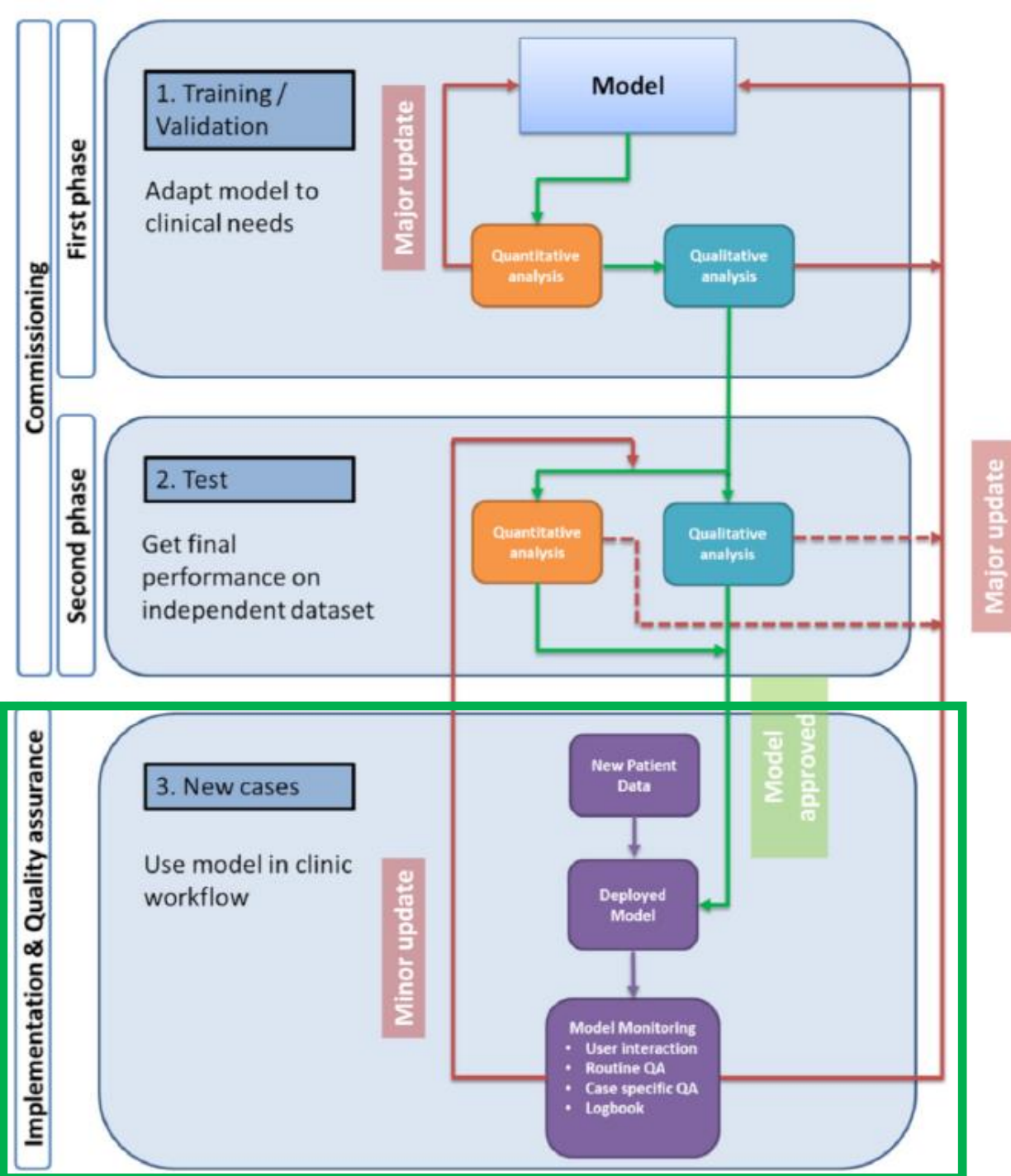
Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." *Radiotherapy and Oncology* 153 (2020): 55-66.



Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." Radiotherapy and Oncology 153 (2020): 55-66.

QA (during clinical use)

Legend





Requirements for Ongoing QA

- Put feedback loop in place for minor / major changes to model
 - Model data may need to be tested & re-trained periodically to remain clinically relevant
 - Decide ahead of time what QA is necessary in each situation
- Case specific QA:
 - Clinician reviews outcome, may adjust results (such as contouring / planning)
 - Best practice: Log modifications to track model performance & provide feedback for future model improvements
- Routine model QA:
 - Quality management program should verify:
 - model has not been corrupted / unintentionally modified
 - remains valid after upgrades / updates
 - Run prediction on a known dataset & compare to prior results
 - Review case specific failures

Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." *Radiotherapy and Oncology* 153 (2020): 55-66.



Summary for Radonc Specific AI Applications

	Automated segmentation	Automated treatment planning	Synthetic CT
Objective(s)	Increase accuracy/consistency Time saving Decrease inter- and intra-observer variability		Using MRI or CBCT for dose calculation (and patient positioning)
Commissioning	<i>training/validation phase</i> : tune model to clinical needs <i>test phase</i> : independent evaluation (accuracy/robustness)		
Data characterization	• Large amount of high quality annotated, retrospective data [2] • Variability of the clinical practice should be represented Consistency in treatment site and/or treatment technique		Consistency in treatment site, image acquisition protocol and image device
Model analysis	<i>Quantitative</i> : calculating similarity to clinical ground truths Overlap, distance and dose measures (cf. Table 1)	Dosimetric measures (cf. Table 2)	Measures to evaluate image similarity, geometric fidelity and dosimetric accuracy (cf. Table 3)
	<i>Qualitative</i> : present output to clinical experts for revision		
Implementation	• Multidisciplinary team with relevant expertise and AI knowledge needed • End-to-end test • Feedback system • Risk analysis (recommended)		
Quality assurance	To monitor the accuracy/consistency of model output over time and robustness to adapted workflows, i.e. continuous quality improvement [28]		
Case-specific QA	• Manual verification • Involvement of all users/education • Statistical or AI models to identify outliers • Independent, secondary automated algorithm or plausibility check		
Routine model QA	• Quality Management Program (QMP) with reference dataset [32] • Informatics platform/log file to collect suboptimal cases, i.e. post-market surveillance [33]		

Vandewinckele, et al. "Overview of AI-based applications in RT: recommendations for implementation & QA." Radiotherapy and Oncology 153 (2020): 55-66.



"A clinical linear accelerator sketched by Leonardo da Vinci."

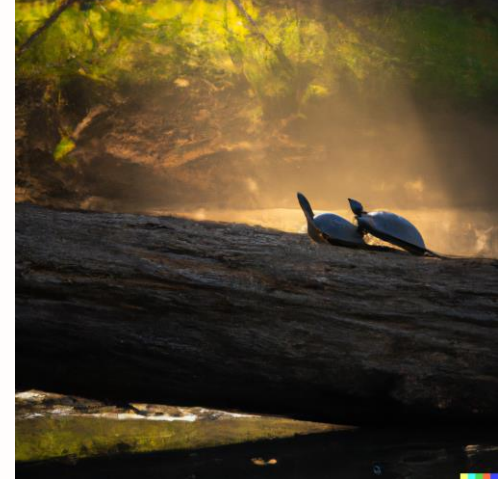


"Steampunk
Radiation Oncology
Clinical Linear
Accelerator."

"A Rembrandt
painting of a
computed
tomography image."



"a national geographic photo of
two turtles sitting on a log in
a river that runs through the
forest during sunrise."



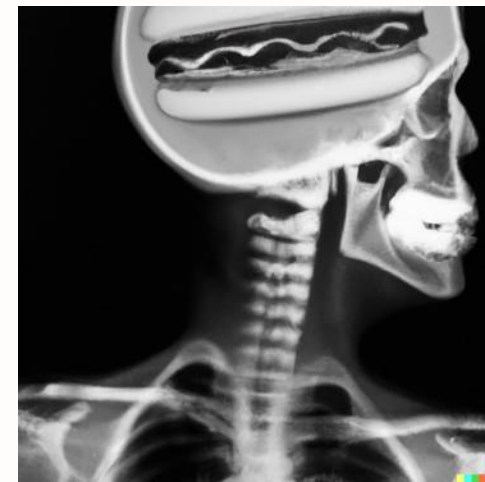
Obrigado!

"Ancient mesoamerican
sculpture of Spongebob
Squarepants."



"A cheeseburger
made out of cotton
and concrete."

"High quality
radiograph of a
head and neck with
a hamburger."



"Steampunk panda
looking at a
radiograph"

