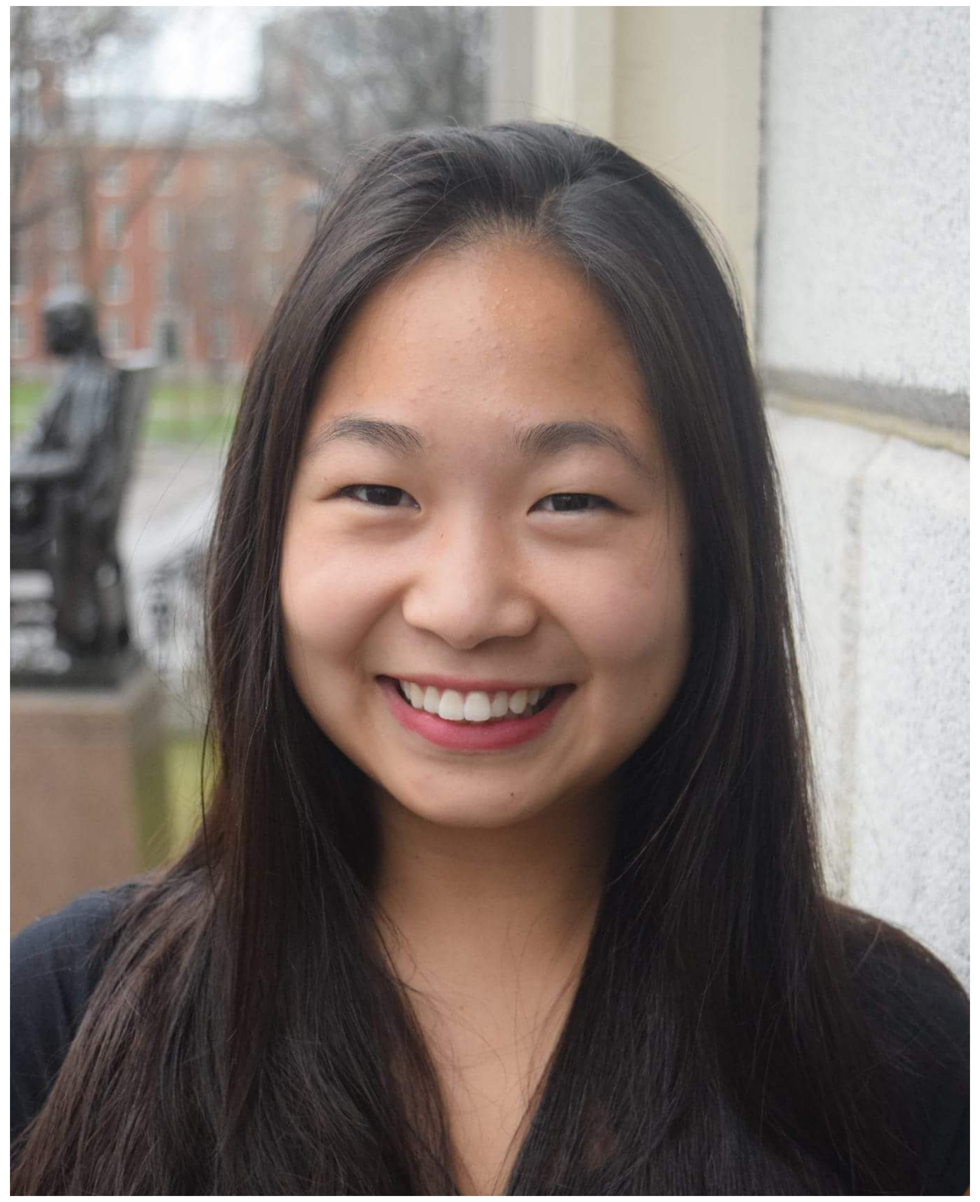
Reimagining the Machine Learning Life Cycle in Education

(and beyond)

Lydia T. Liu | BAIR/CPAR/BDD talk, Feb 10 2022

Joint work with co-authors



Lydia T. Liu*
UC Berkeley EECS

Serena Wang*
UC Berkeley EECS

Rediet Abebe†
UC Berkeley EECS

Tolani Britton†
UC Berkeley GSE

* Equal authorship †Equal advisorship

Machine Learning (ML) for Social Good?

AI for Social Good

Machine Learning (ML) for Social Good?

- Significant, growing research agenda on “ML/AI for social good”

Machine Learning (ML) for Social Good?

- Significant, growing research agenda on “ML/AI for social good”
  - Multiple academic workshops

AI for Social Good

Machine Learning (ML) for Social Good?

- Significant, growing research agenda on “ML/AI for social good”
  - Multiple academic workshops
  - Applications in education, public health, environmental protection, social services, and more.

Machine Learning (ML) for Social Good?

- Significant, growing research agenda on “ML/AI for social good”
  - Multiple academic workshops
  - Applications in education, public health, environmental protection, social services, and more.
- **Two prevailing challenges**

AI for Social Good

Machine Learning (ML) for Social Good?

- Significant, growing research agenda on “ML/AI for social good”
  - Multiple academic workshops
  - Applications in education, public health, environmental protection, social services, and more.
- **Two prevailing challenges**
  1. Limited evidence of long-term effectiveness.

Machine Learning (ML) for Social Good?

- Significant, growing research agenda on “ML/AI for social good”
  - Multiple academic workshops
  - Applications in education, public health, environmental protection, social services, and more.
- **Two prevailing challenges**
  1. Limited evidence of long-term effectiveness.
  2. Limited inquiry into what “social good” entails, and whether and how ML4SG efforts contribute.

Machine Learning (ML) in Education Today

Machine Learning (ML) in Education Today

- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.

Machine Learning (ML) in Education Today

- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.

Machine Learning (ML) in Education Today

- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.
  - Predicting student dropout risk

edX coursera

Machine Learning (ML) in Education Today

- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.
  - Predicting student dropout risk
  - Automated grading/essay scoring

Source: Nancy E Bailey

Machine Learning (ML) in Education Today

- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.
  - Predicting student dropout risk
  - Automated grading/essay scoring
  - Admissions

edX coursera

Source: Nancy E Bailey

Machine Learning (ML) in Education Today

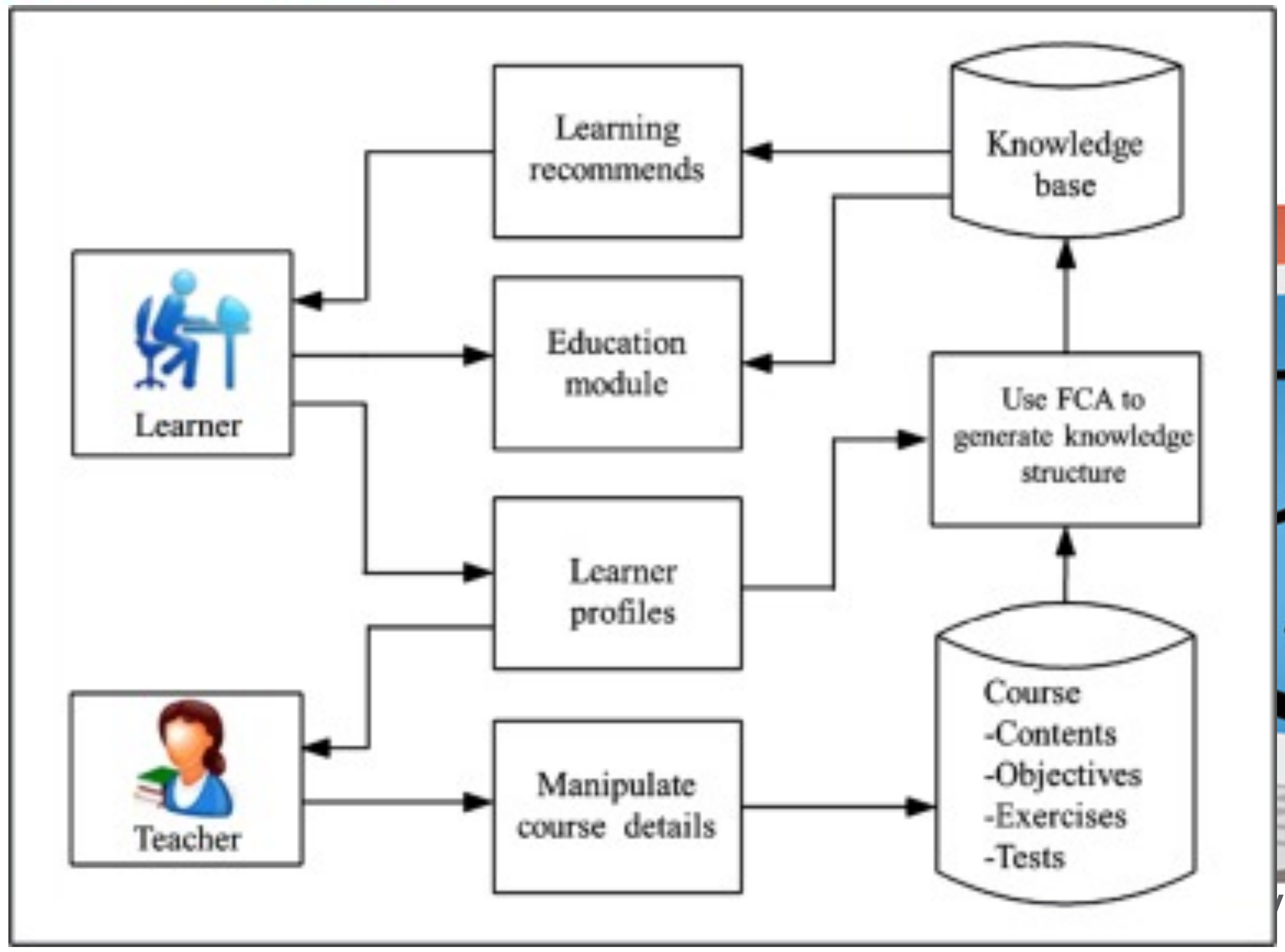
- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.
    - Predicting student dropout risk
    - Automated grading/essay scoring
    - Admissions
    - Predicting test question difficulty

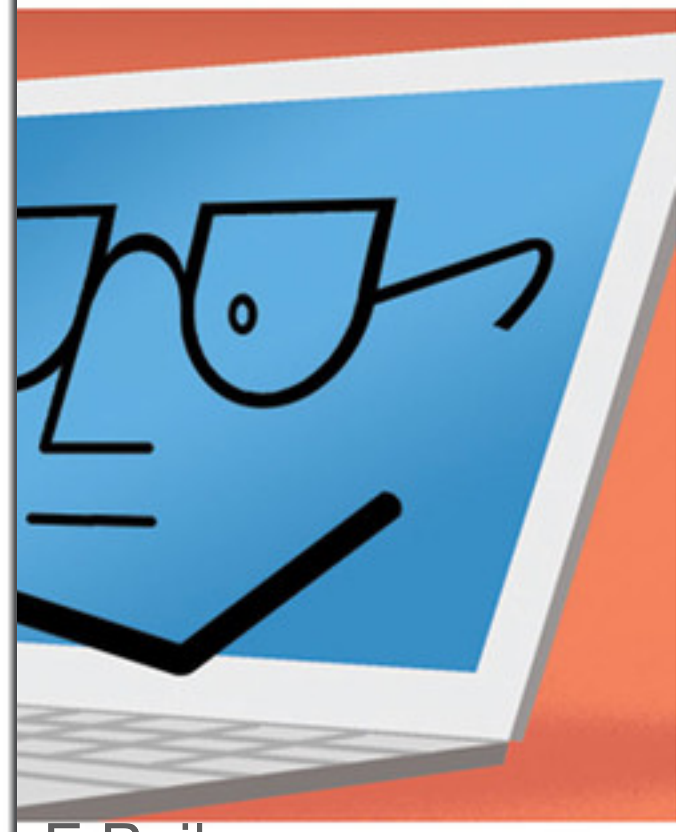
edX coursera

Source: Nancy E Bailey

Machine Learning (ML) in Education Today

- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.
    - Predicting student dropout risk
    - Automated grading/essay scoring
    - Admissions
    - Predicting test question difficulty
    - Intelligent tutoring systems

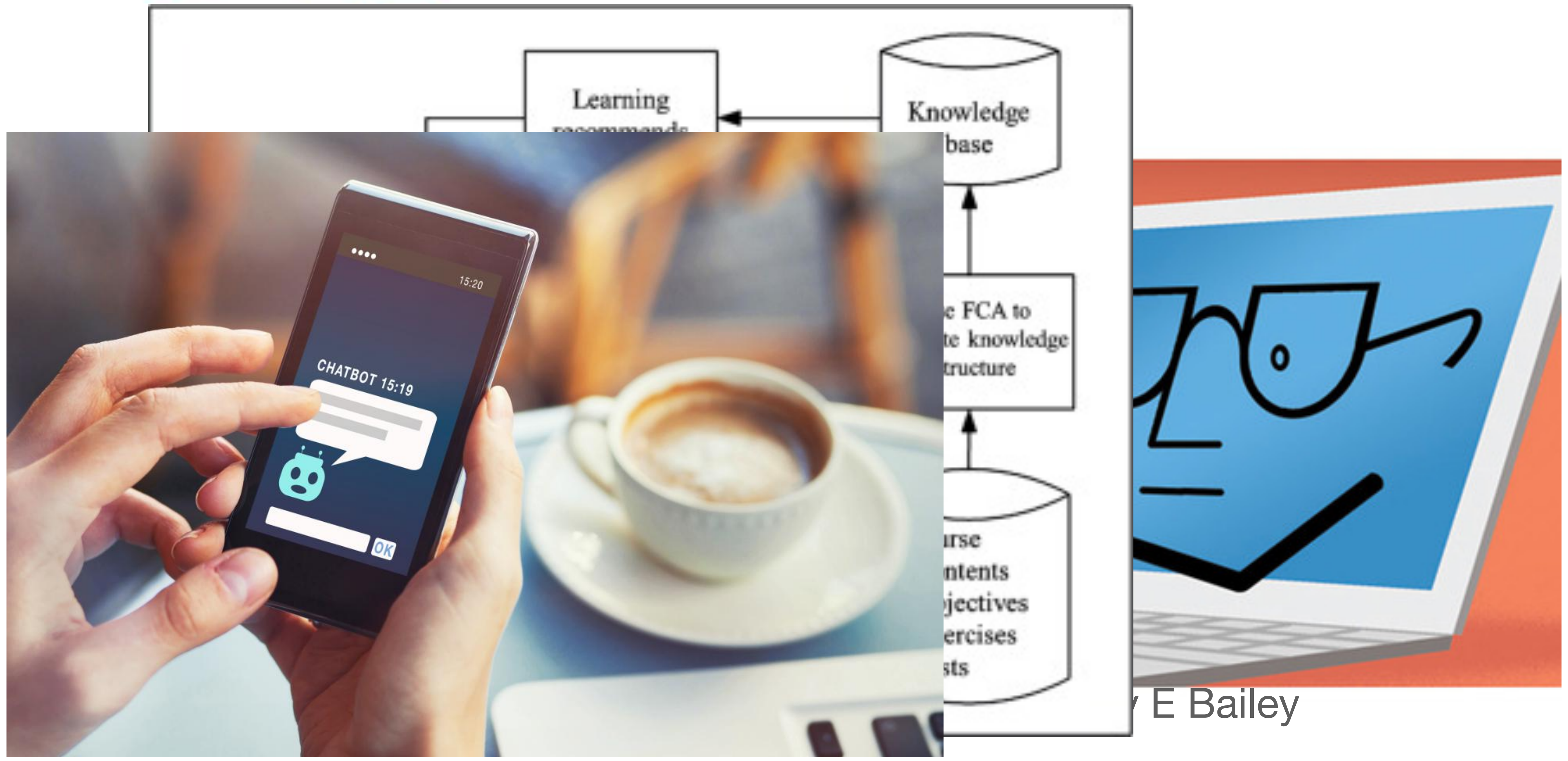




E Bailey

Machine Learning (ML) in Education Today

- Machine learning (ML) techniques are prevalent in the education sphere, not only in Massive Open Online Courses (MOOCs), but also in higher education and K-12.
  - Predicting student dropout risk
  - Automated grading/essay scoring
  - Admissions
  - Predicting test question difficulty
  - Intelligent tutoring systems
  - Chat bots for advising



Recent controversies

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Inputs:
GPA, GRE, letters of recommendation, personal statements

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Inputs:
GPA, GRE, letters of recommendation, personal statements

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Inputs:
GPA, GRE, letters of recommendation, personal statements

→

Target:
Past admission outcome (2009-2012)

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Inputs:
GPA, GRE, letters of recommendation, personal statements

→

Target:
Past admission outcome (2009-2012)

- Usage: if GRADE predicts low enough, applicant may be rejected without further review.

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Inputs:
GPA, GRE, letters of recommendation, personal statements

→

Target:
Past admission outcome (2009-2012)

- Usage: if GRADE predicts low enough, applicant may be rejected without further review.

The Death and Life of an Admissions Algorithm

Source: InsideHigherEd

> U of Texas at Austin has stopped using a machine-learning system to evaluate applicants for its Ph.D. in computer science. Critics say the system exacerbates existing inequality in the field.

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Computer Science at UT Austin
@UTCompSci

Replying to @yasmmeme

TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which is why we have phased out our reliance on GRADE and are no longer using it as part of our graduate admissions process.

1:46 PM · Dec 1, 2020 · Twitter Web App

Source: Twitter

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Computer Science at UT Austin
@UTCompSci

Replying to @yasmmeme

TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which is why we have phased out our reliance on GRADE and are no longer using it as part of our graduate admissions process.

1:46 PM · Dec 1, 2020 · Twitter Web App

Source: Twitter

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Computer Science at UT Austin
@UTCompSci

Replying to @yasmmeme

TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which is why we have phased out our reliance on GRADE and are no longer using it as part of our graduate admissions process.

1:46 PM · Dec 1, 2020 · Twitter Web App

“social good”

Source: Twitter

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

Computer Science at UT Austin
@UTCompSci

Replying to @yasmmeme

TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which is why we have phased out our reliance on GRADE and are no longer using it as part of our graduate admissions process.

1:46 PM · Dec 1, 2020 · Twitter Web App

“social good”

Source: Twitter

Recent controversies

- GRADE algorithm for graduate admissions at UT Austin

TEXAS
Computer Science

Computer Science at UT Austin
@UTCompSci

Replying to @yasmmeme

“social good”

TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which is why we have phased out our reliance on GRADE and are no longer using it as part of our graduate admissions process.

How the ML system “falls short”…?

1:46 PM · Dec 1, 2020 · Twitter Web App

Source: Twitter

Question: Are the stated or implied "social good" objectives of ML4Ed research papers aligned with the ML tasks, objectives, and datasets? Why (not)?

Research methodology

Research methodology

ML for education papers:

Research methodology

ML for education papers:

Education researchers:

Research methodology

ML for education papers:

Education researchers:

Research methodology

ML for education papers:

Education researchers:

- 15 interview participants recruited for expertise (e.g. faculty in graduate schools of education)

Research methodology

ML for education papers:

- 20 papers sampled from Shi et al. (2020)'s survey on AI for social good. Published in ML conferences such as AAAI and KDD.

Education researchers:

- 15 interview participants recruited for expertise (e.g. faculty in graduate schools of education)

Research methodology

ML for education papers:

- 20 papers sampled from Shi et al. (2020)'s survey on AI for social good. Published in ML conferences such as AAAI and KDD.

Education researchers:

- 15 interview participants recruited for expertise (e.g. faculty in graduate schools of education)

- Major education topics covered by papers: higher education, student learning, MOOCs, and standardized assessment.

Research methodology

ML for education papers:

- 20 papers sampled from Shi et al. (2020)'s survey on AI for social good. Published in ML conferences such as AAAI and KDD.

Education researchers:

- 15 interview participants recruited for expertise (e.g. faculty in graduate schools of education)

- Major education topics covered by papers: higher education, student learning, MOOCs, and standardized assessment.
- Notable omissions: special education, early education, teaching

Research methodology

Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)

Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.

Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.
  - **Part I:** participant's background in education.
    Part II: a focused discussion on the research paper.
    E.g. "How would you describe their goals?", and "To what extent do you agree this machine learning task captures the ... goal?"

Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.
  - **Part I:** participant's background in education.
    Part II: a focused discussion on the research paper.
    E.g. "How would you describe their goals?", and "To what extent do you agree this machine learning task captures the ... goal?"
- Data analysis: inductive process

Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.
  - **Part I:** participant's background in education.
    Part II: a focused discussion on the research paper.
    E.g. "How would you describe their goals?", and "To what extent do you agree this machine learning task captures the ... goal?"
- Data analysis: inductive process

Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.
  - **Part I:** participant's background in education.
    Part II: a focused discussion on the research paper.
    E.g. "How would you describe their goals?", and "To what extent do you agree this machine learning task captures the ... goal?"
- Data analysis: inductive process

Start with interview transcript data

Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.
  - **Part I:** participant's background in education.
    Part II: a focused discussion on the research paper.
    E.g. "How would you describe their goals?", and "To what extent do you agree this machine learning task captures the ... goal?"
- Data analysis: inductive process

Start with interview transcript data

Group into themes

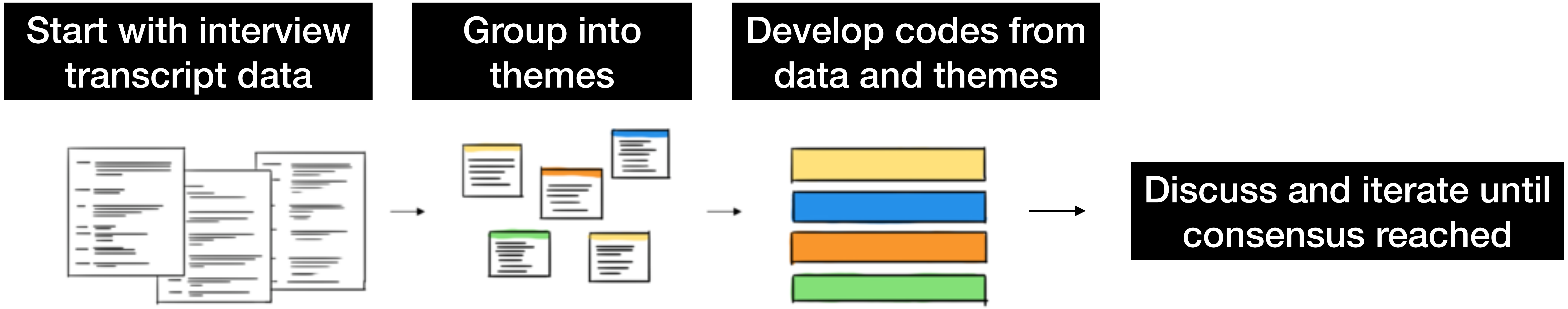
Research methodology

- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.
  - **Part I:** participant’s background in education.
    Part II: a focused discussion on the research paper.
    E.g. “How would you describe their goals?”, and “To what extent do you agree this machine learning task captures the ... goal?”
- Data analysis: inductive process

Start with interview transcript data | Group into themes | Develop codes from data and themes

Research methodology

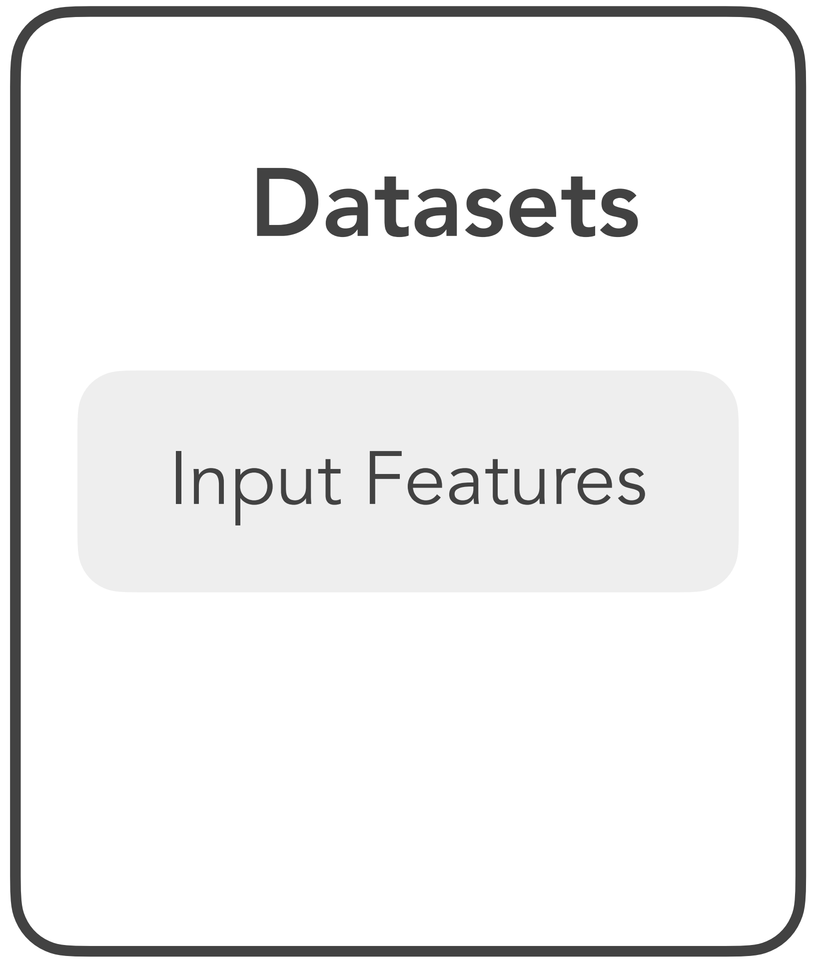
- Interviews (50-60 minutes each, December 2020 - September 2021)
  - Each interview participant matched to one paper by participant preference.
  - **Part I:** participant's background in education.
    Part II: a focused discussion on the research paper.
    E.g. "How would you describe their goals?", and "To what extent do you agree this machine learning task captures the ... goal?"
- Data analysis: inductive process



(Supervised) Machine Learning Paradigm

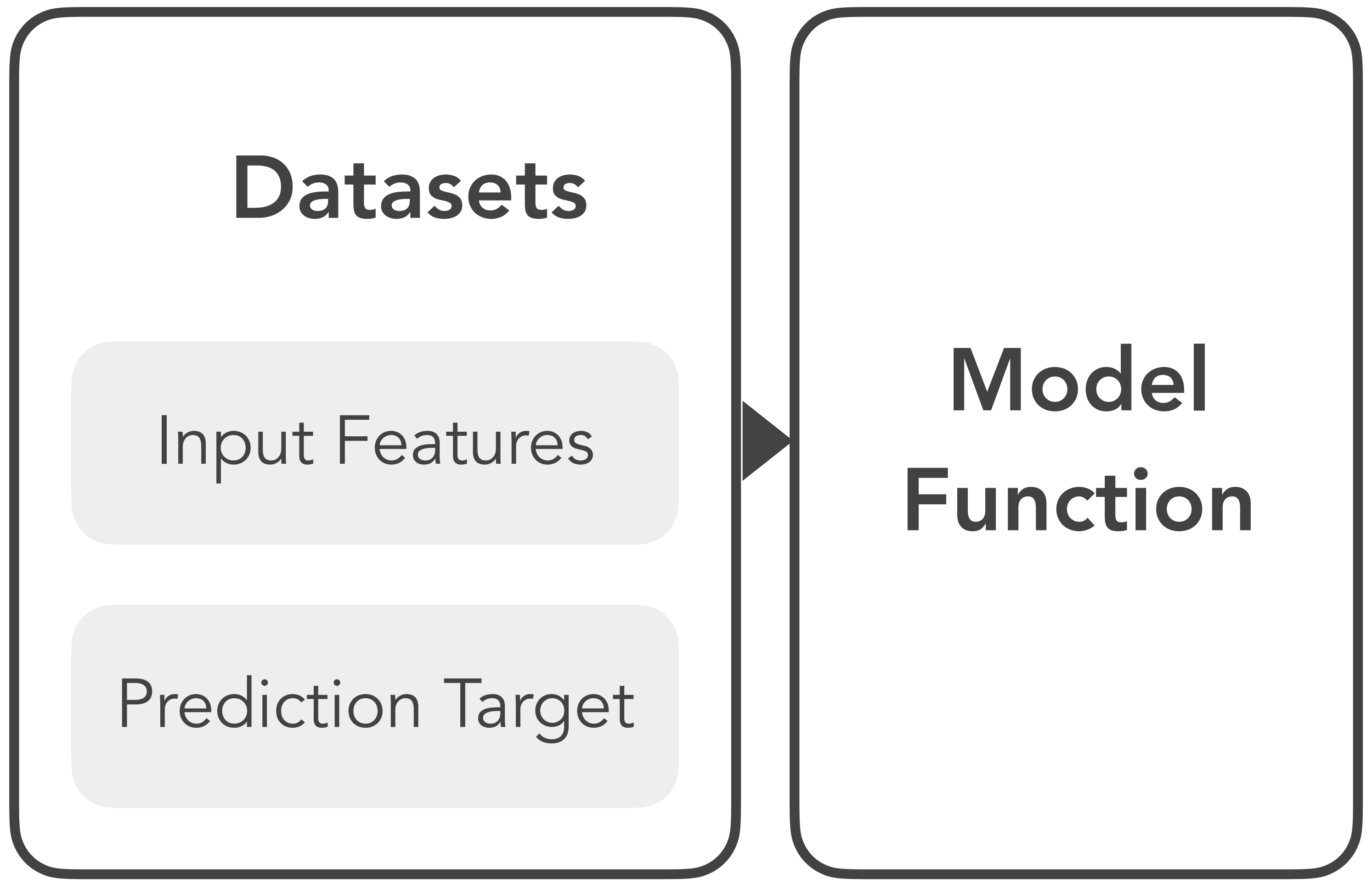
(Supervised) Machine Learning Paradigm

(Supervised) Machine Learning Paradigm

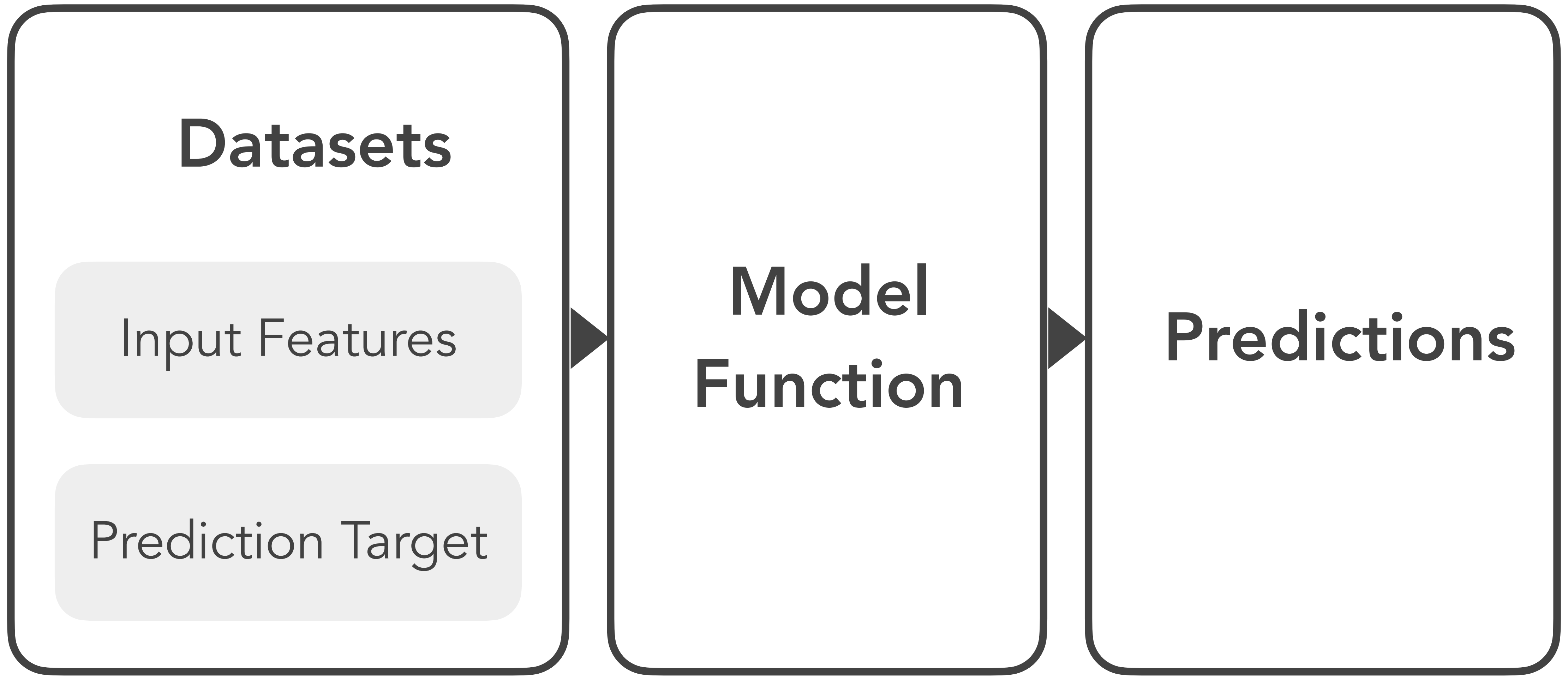


(Supervised) Machine Learning Paradigm

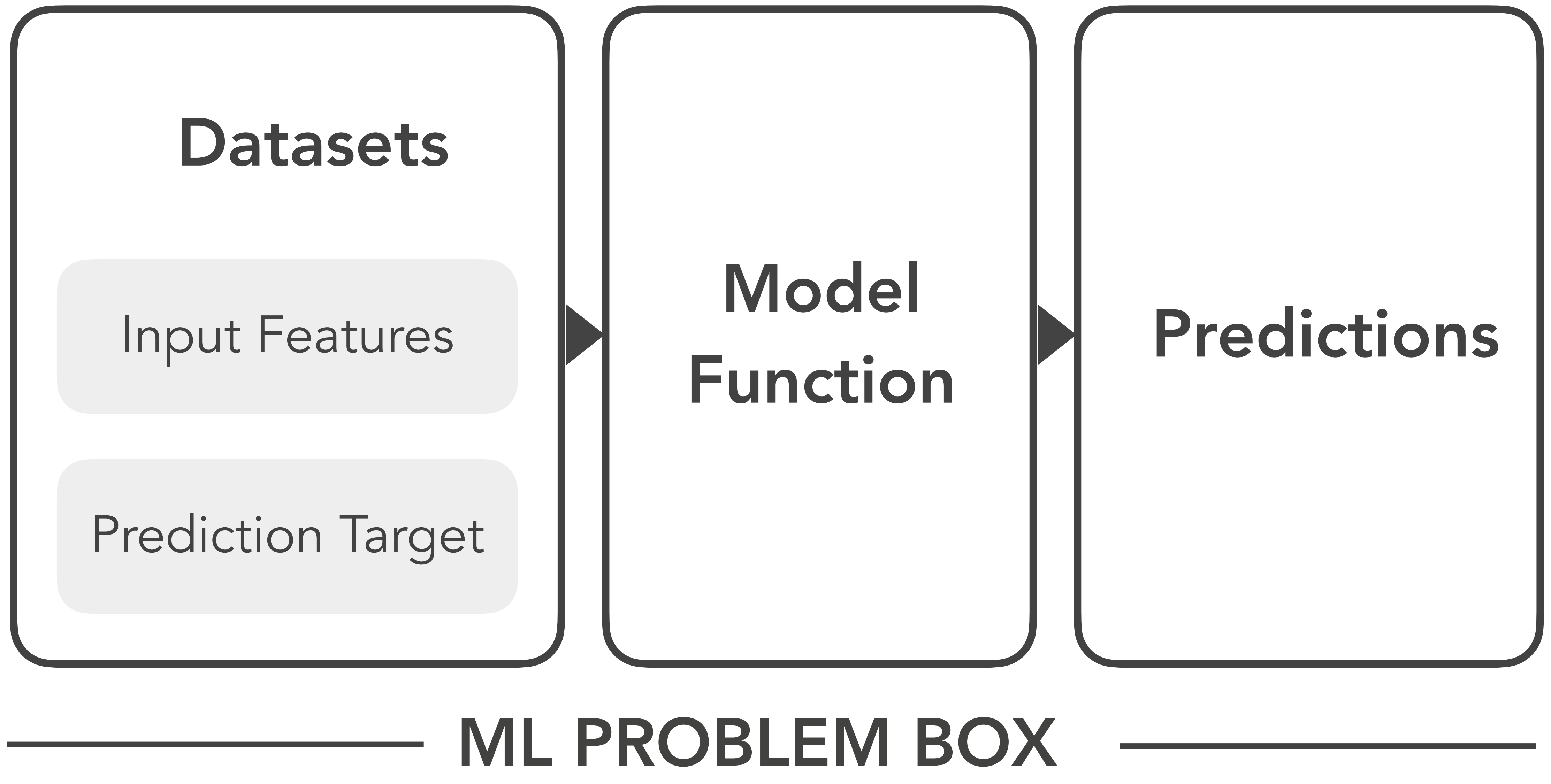
(Supervised) Machine Learning Paradigm



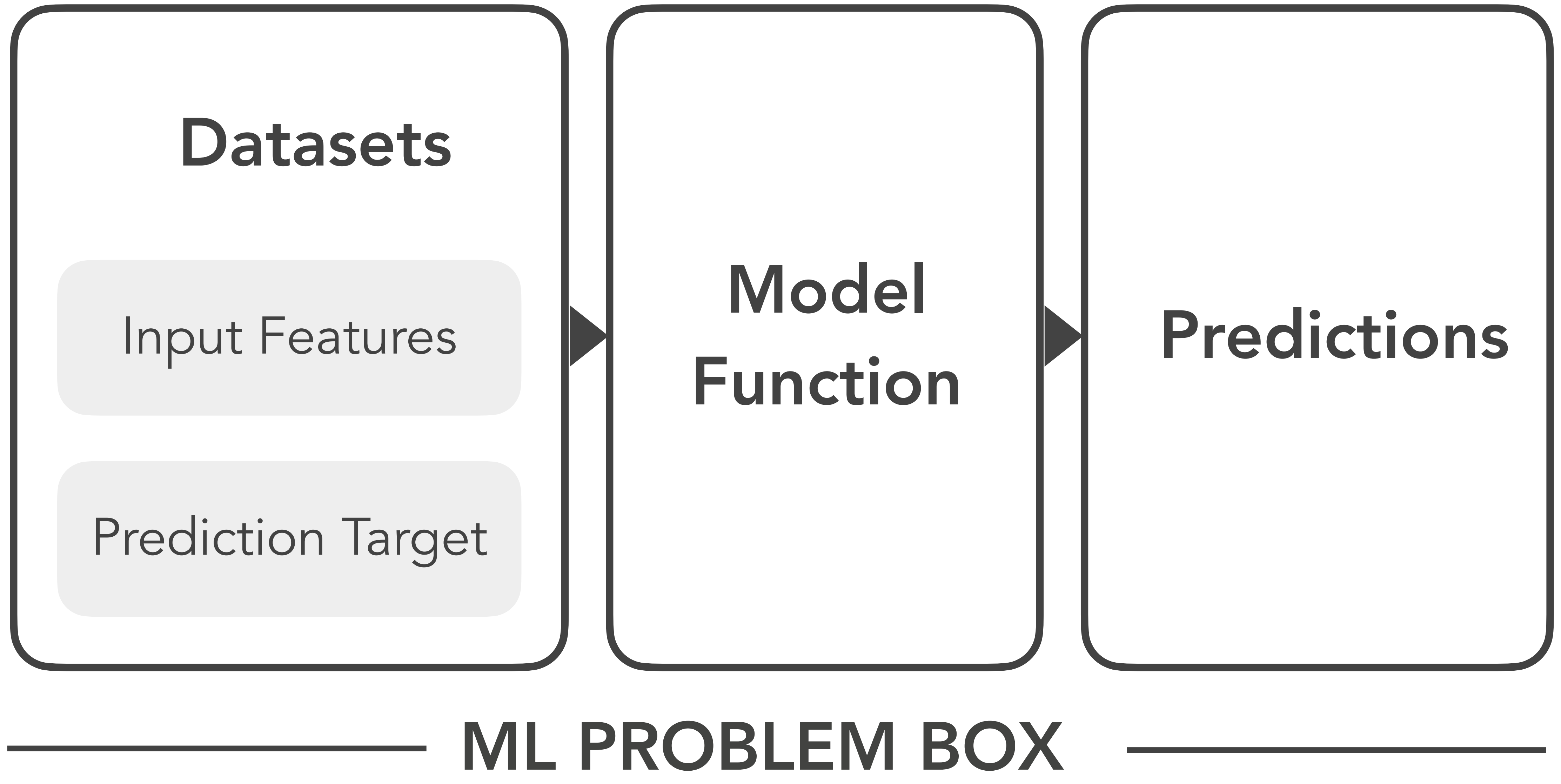
(Supervised) Machine Learning Paradigm

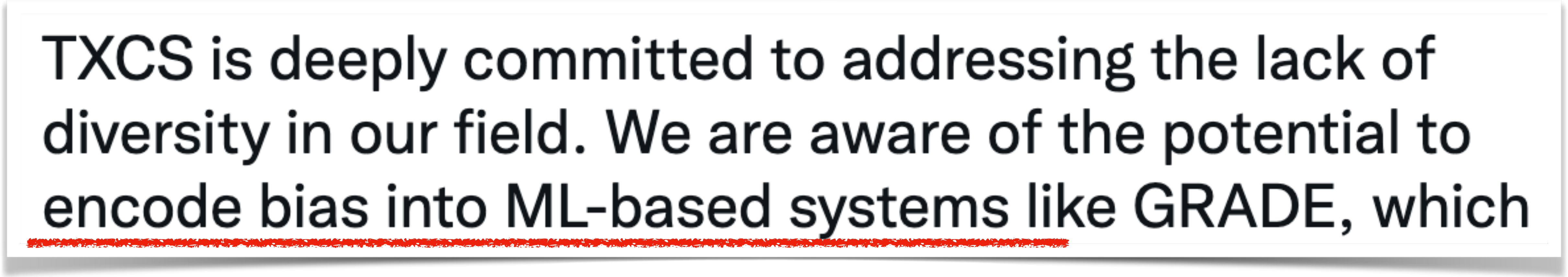


(Supervised) Machine Learning Paradigm



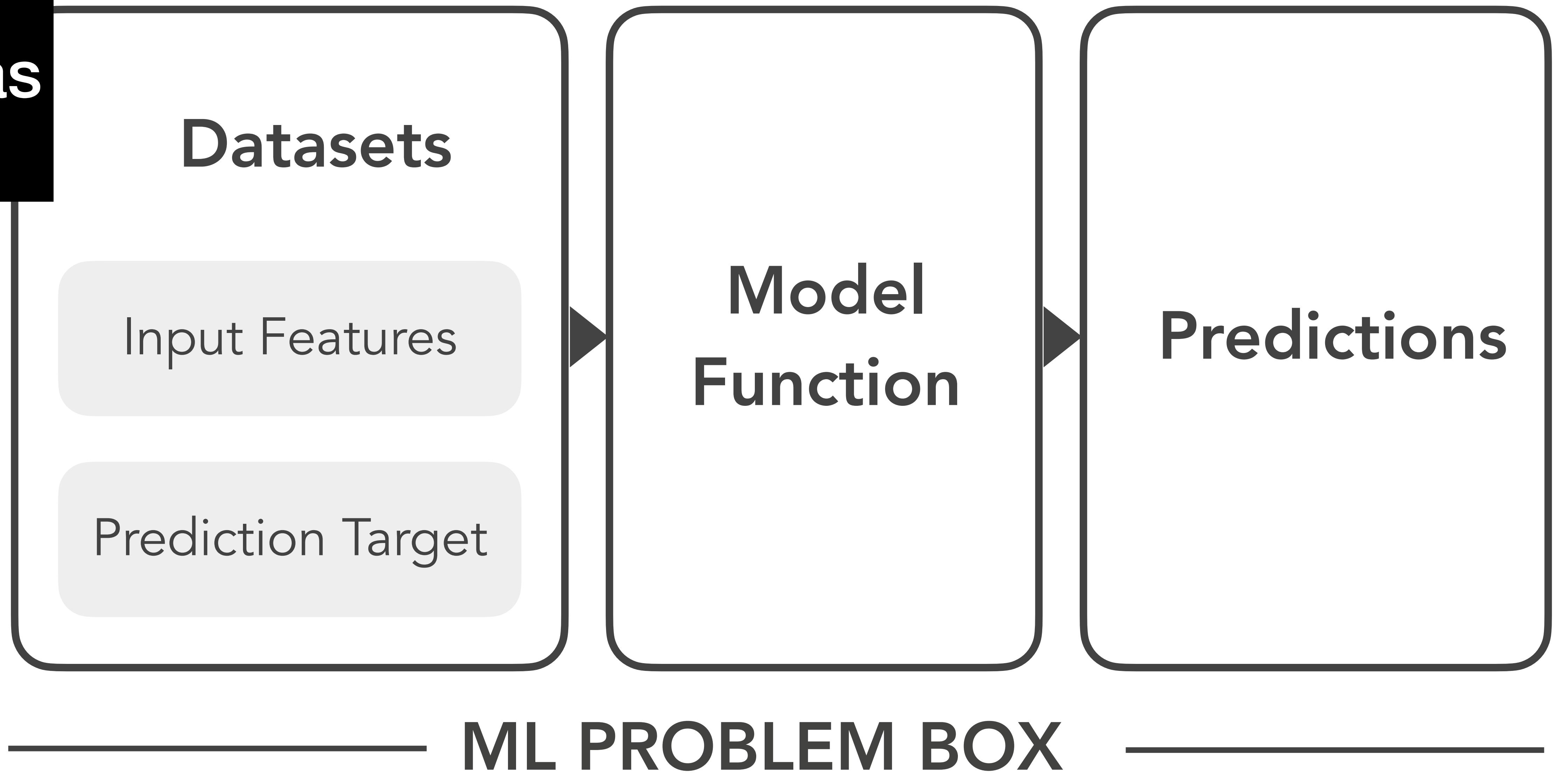
(Supervised) Machine Learning Paradigm




TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which

(Supervised) Machine Learning Paradigm

Addressing “bias” in data: learning fair representations (e.g. Zemel et al 2013), label bias (e.g. Obermeyer et al 2019)

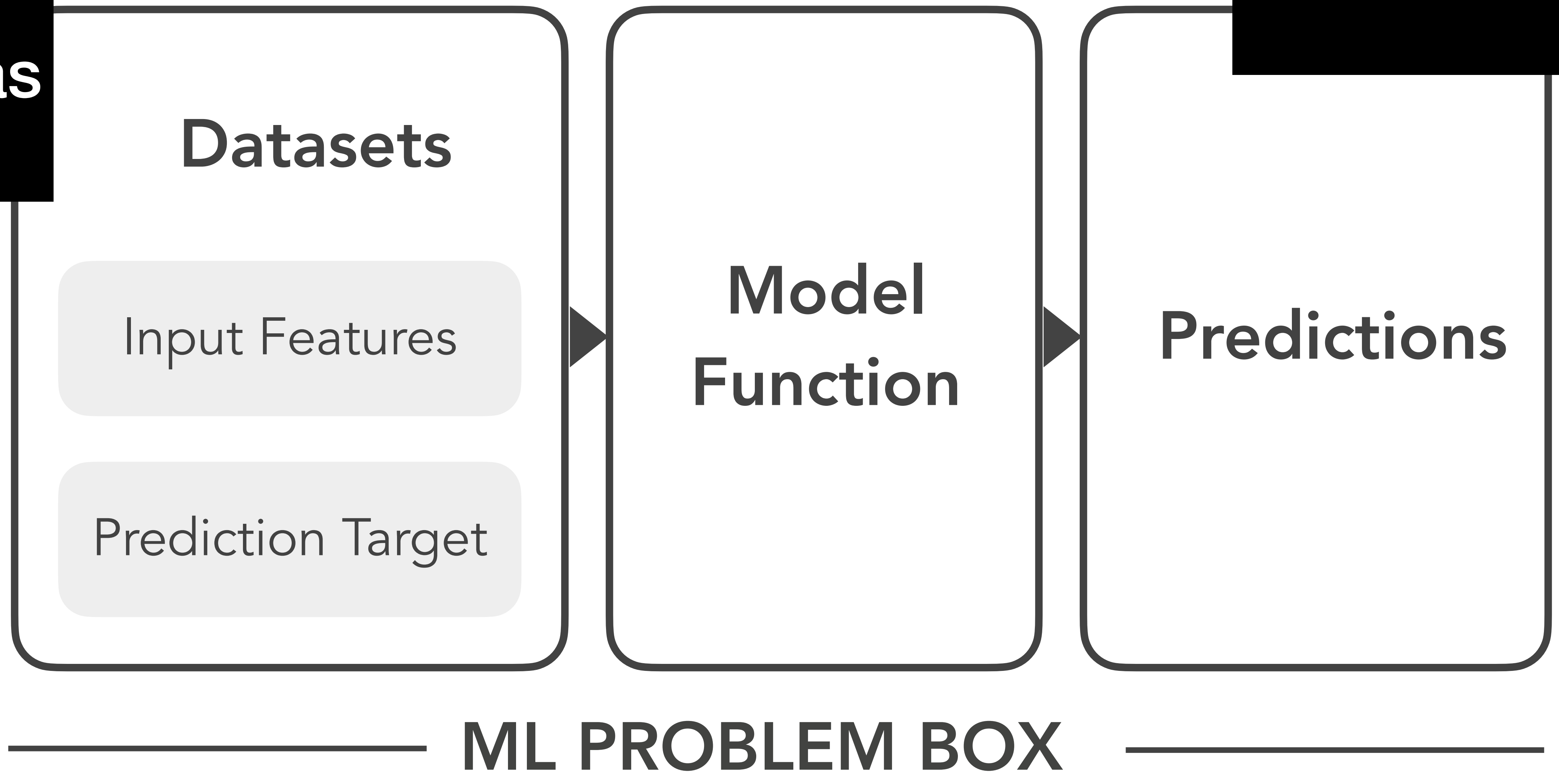


TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which

(Supervised) Machine Learning Paradigm

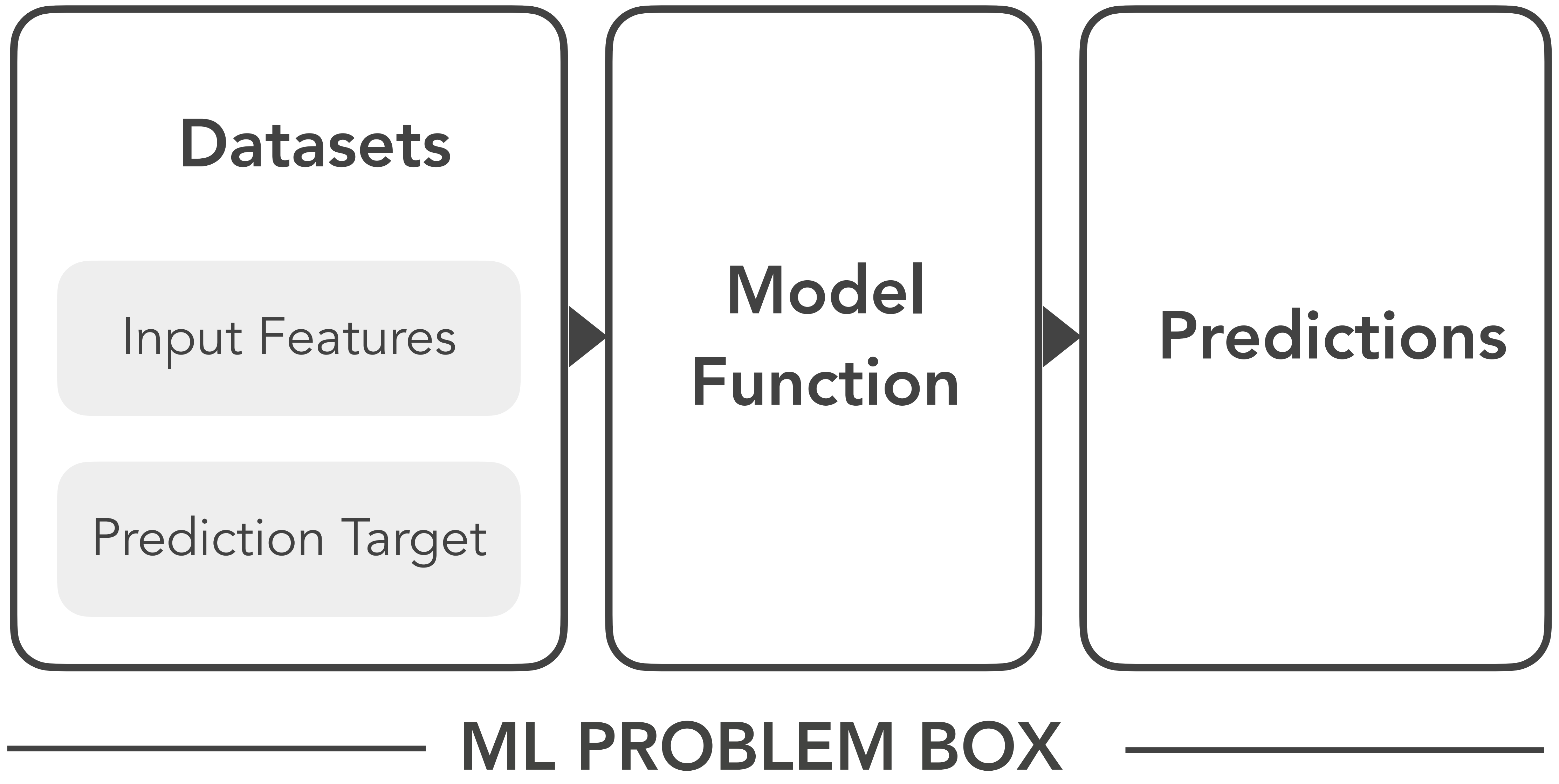
Addressing “bias” in data: learning fair representations (e.g. Zemel et al 2013), label bias (e.g. Obermeyer et al 2019)

Addressing “bias” in prediction: algorithmic fairness in ML (e.g. Dwork et al 2011, Hardt et al 2016)

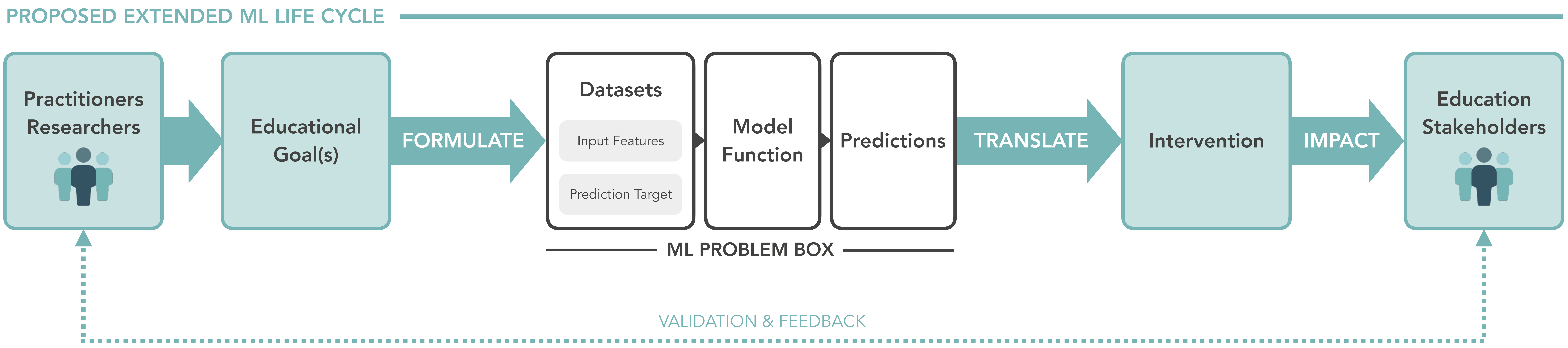


TXCS is deeply committed to addressing the lack of diversity in our field. We are aware of the potential to encode bias into ML-based systems like GRADE, which

Reimagining the ML life cycle

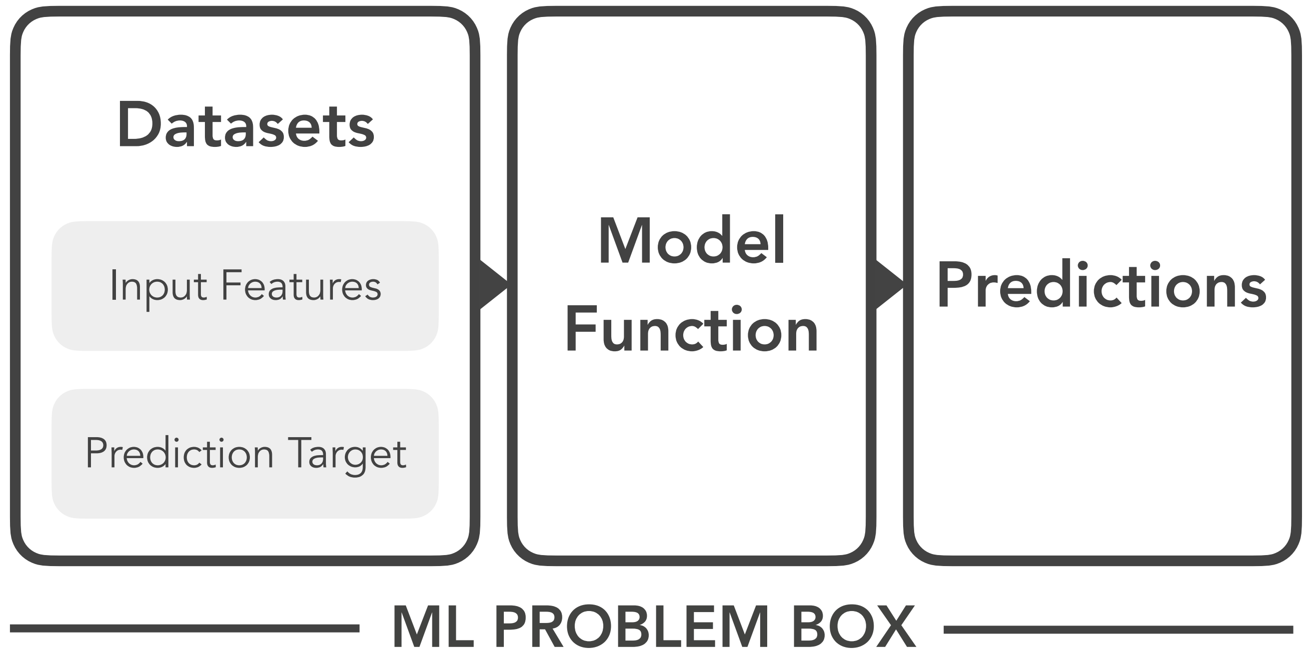


Reimagining the ML life cycle



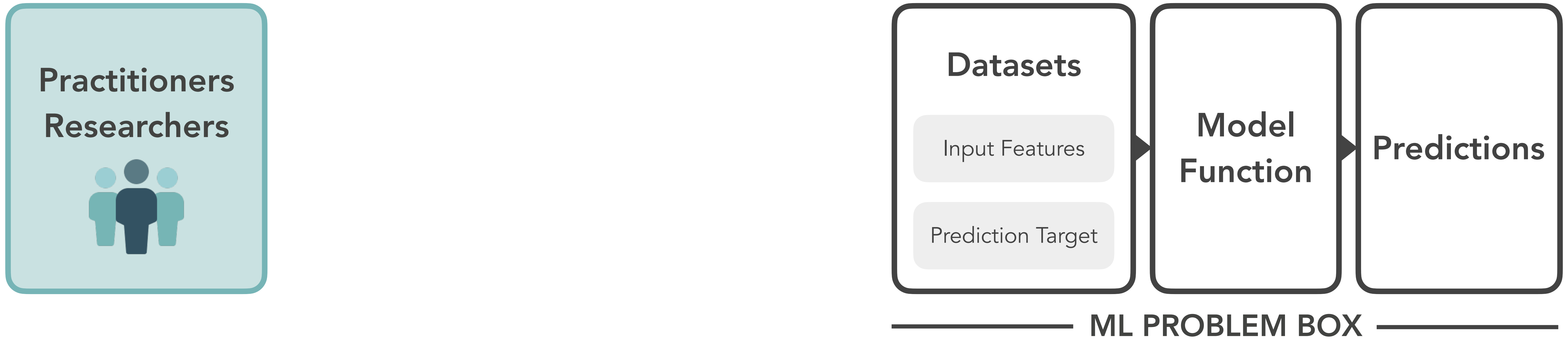
Reimagining the ML life cycle

PROPOSED EXTENDED ML LIFE CYCLE



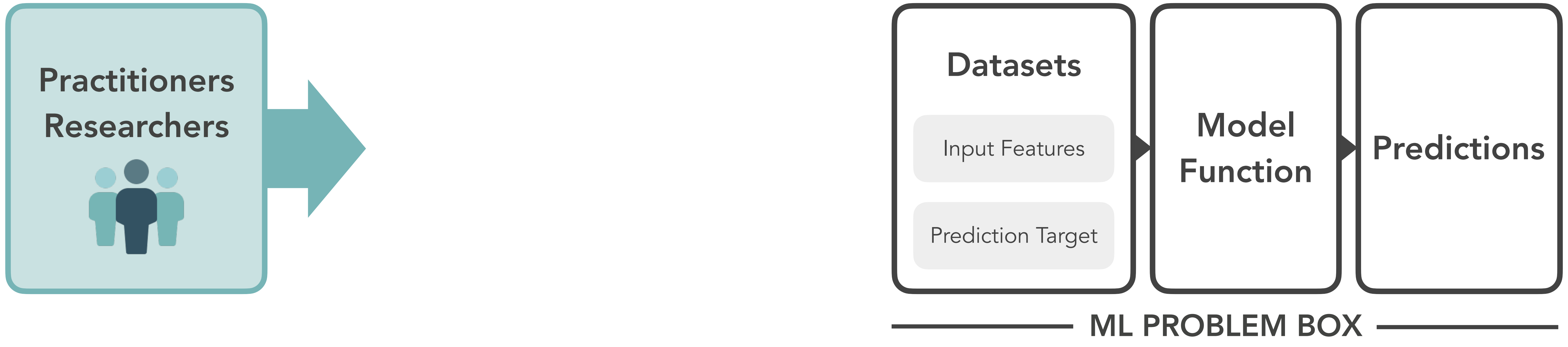
Reimagining the ML life cycle

PROPOSED EXTENDED ML LIFE CYCLE

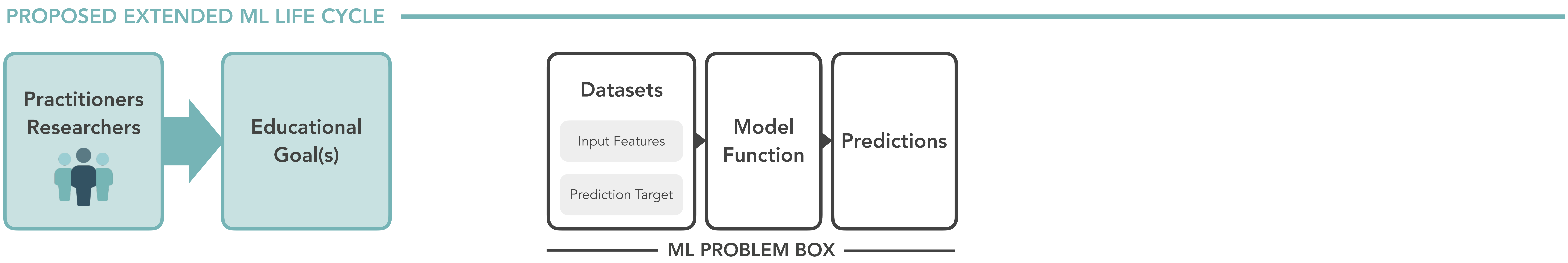


Reimagining the ML life cycle

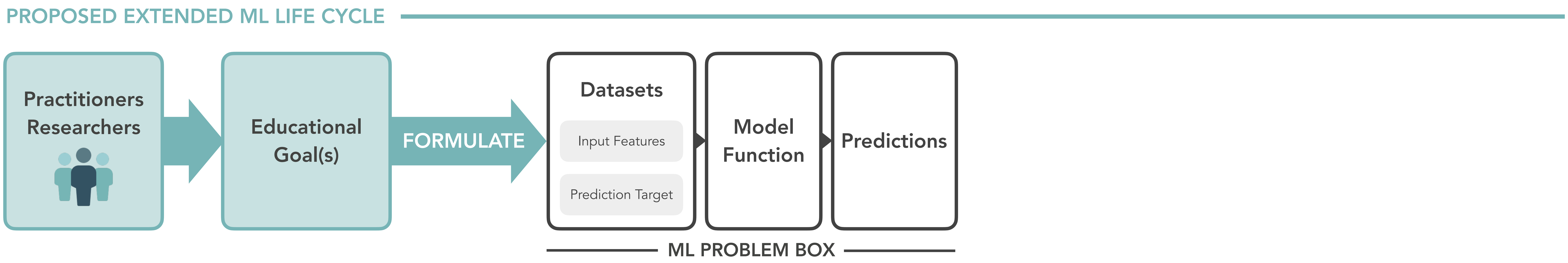
PROPOSED EXTENDED ML LIFE CYCLE



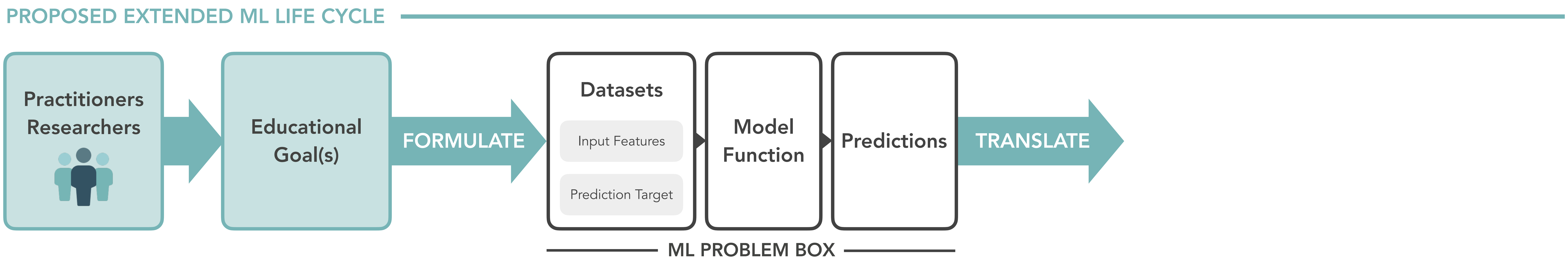
Reimagining the ML life cycle



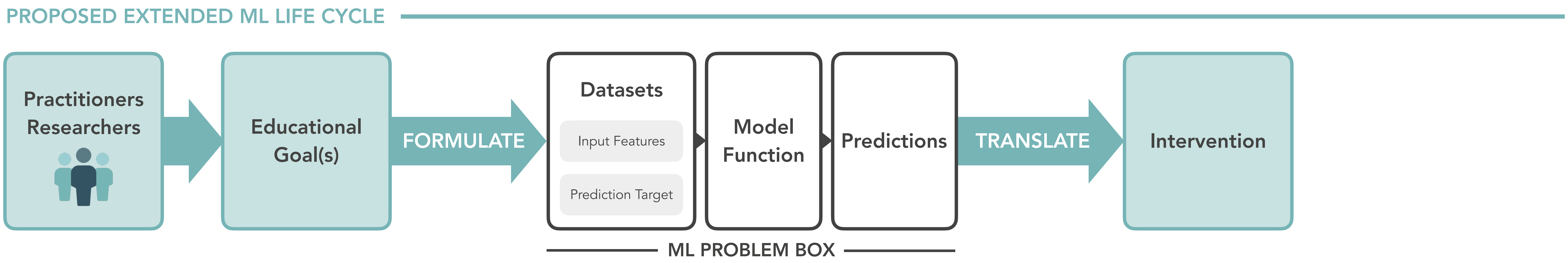
Reimagining the ML life cycle



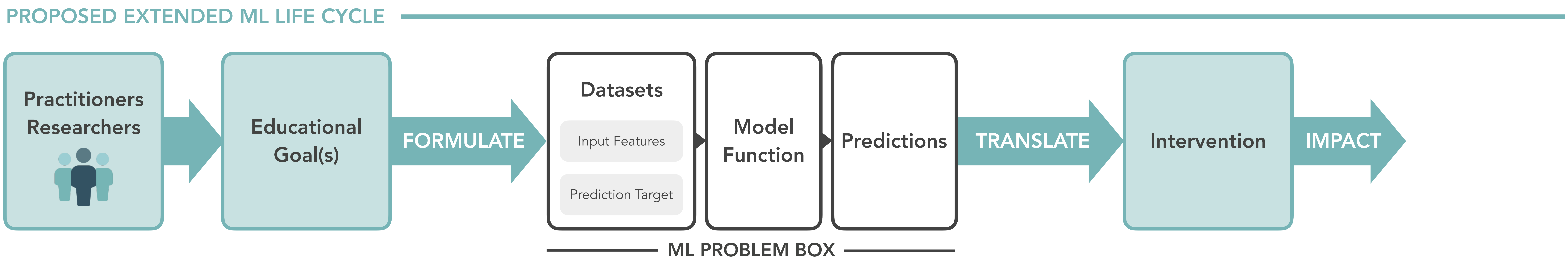
Reimagining the ML life cycle



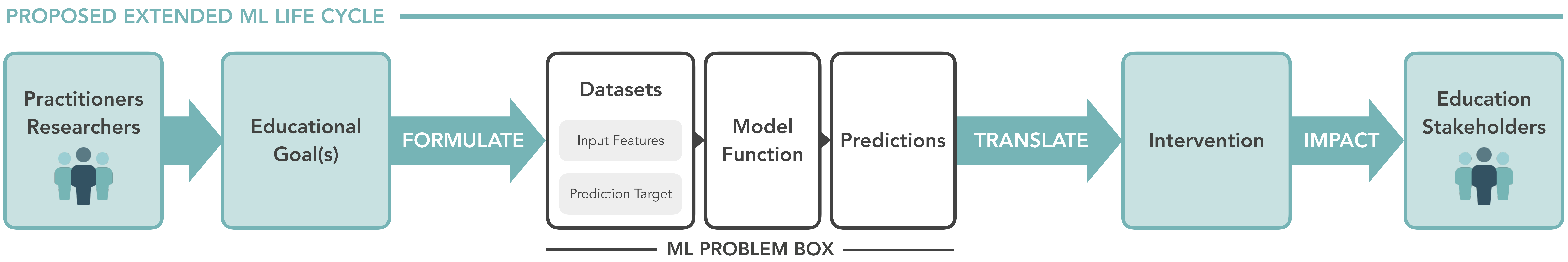
Reimagining the ML life cycle



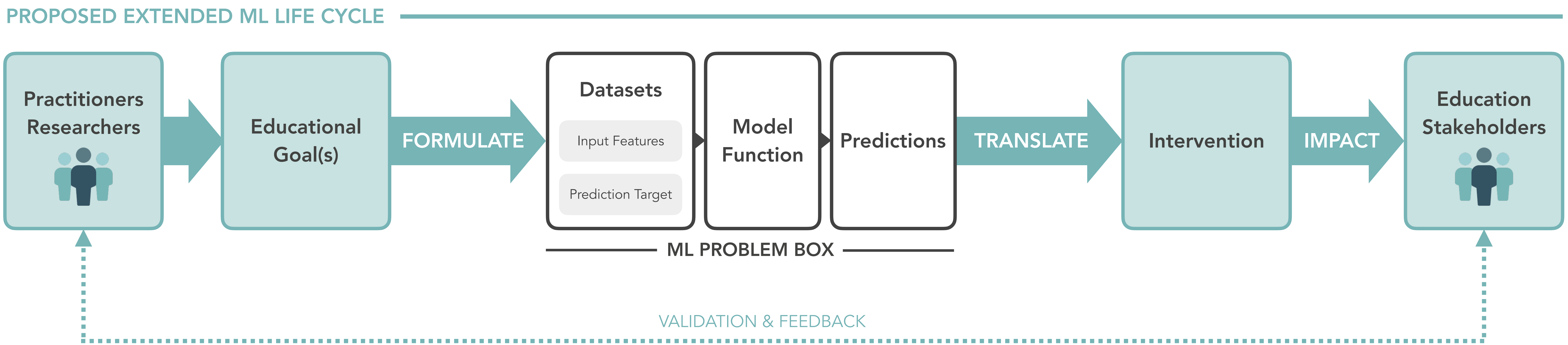
Reimagining the ML life cycle



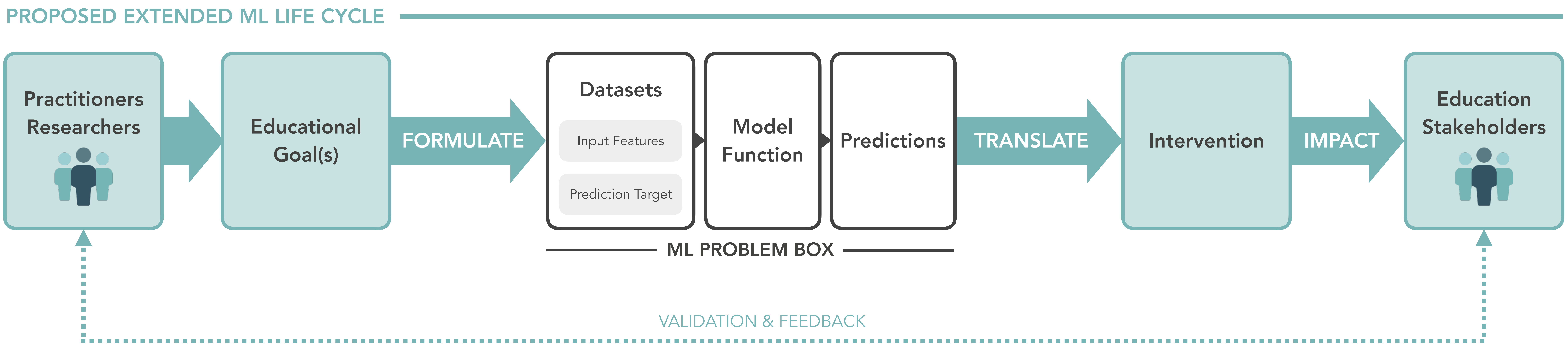
Reimagining the ML life cycle



Reimagining the ML life cycle

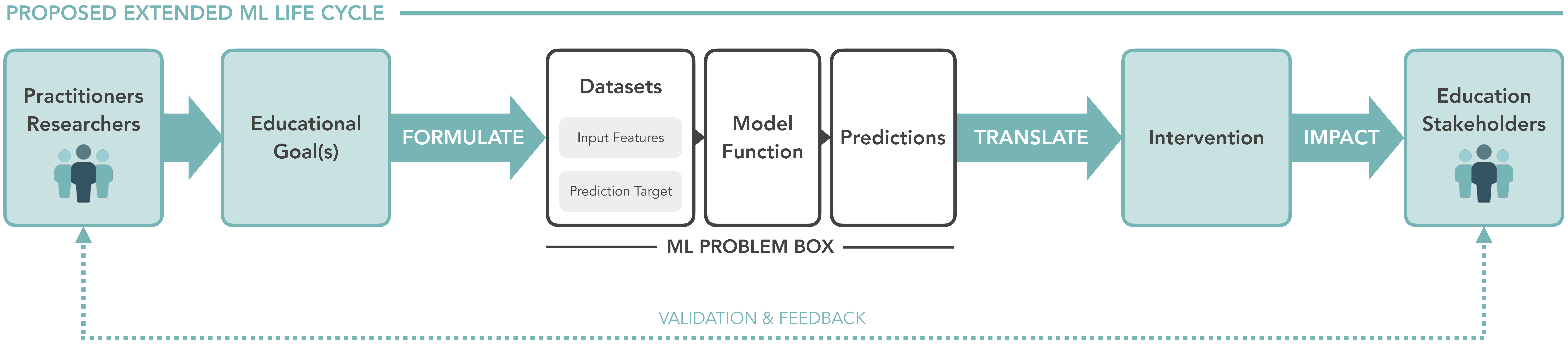


Reimagining the ML life cycle



Part 1: Translating Education Goals via Problem Formulation

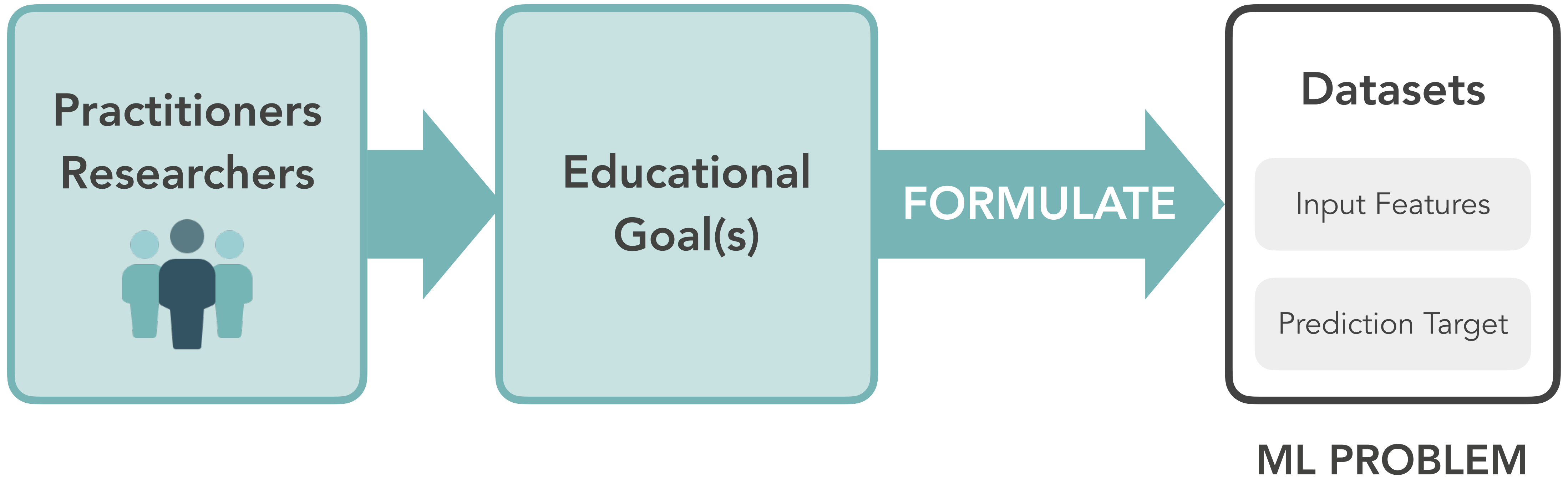
Reimagining the ML life cycle



Part 1: Translating Education Goals via Problem Formulation

Part 2: Translating Predictions to Interventions

Part 1: Translating Education Goals via Problem Formulation



Conflating education goals with short-term metrics

Conflating education goals with short-term metrics

- ML4Ed papers focus on a **single quantitative metric**, e.g. “improving on-time graduation rates” where the actual education goal is broader and longer-term.

Conflating education goals with short-term metrics

- ML4Ed papers focus on a **single quantitative metric**, e.g. "improving on-time graduation rates" where the actual education goal is broader and longer-term.

> *P1: "We don't limit ourselves to graduation rates. We're also looking at other things such as retention rates and so forth."*

Conflating education goals with short-term metrics

- ML4Ed papers focus on a **single quantitative metric**, e.g. “improving on-time graduation rates” where the actual education goal is broader and longer-term.

P1: “We don't limit ourselves to graduation rates. We're also looking at other things such as retention rates and so forth.”

P2: “Graduating with satisfaction or knowing exactly what they want to do afterwards. That's another thing that we could look at.”

Conflating education goals with short-term metrics

- ML4Ed papers focus on a **single quantitative metric**, e.g. "improving on-time graduation rates" where the actual education goal is broader and longer-term.

P1: "We don't limit ourselves to graduation rates. We're also looking at other things such as retention rates and so forth."

P2: "Graduating with satisfaction or knowing exactly what they want to do afterwards. That's another thing that we could look at."

*P3: "There is often **bias towards shorter term outcomes** without drawing out the logical map of why do we care" partly because "there is better data about them [...] they're more often in the same dataset".*

Conflating education goals with short-term metrics

Conflating education goals with short-term metrics

- Education institutions can have competing goals, e.g. in admissions. These tensions tend not to be acknowledged in ML4Ed projects.

Conflating education goals with short-term metrics

- Education institutions can have competing goals, e.g. in admissions. These tensions tend not to be acknowledged in ML4Ed projects.

> ***Efficiency goal**: Universities "want students who would have higher rates of retention or graduation […] otherwise, that's a financial loss to them". (P6)*

Conflating education goals with short-term metrics

- Education institutions can have competing goals, e.g. in admissions. These tensions tend not to be acknowledged in ML4Ed projects.

***Efficiency goal**: Universities "want students who would have higher rates of retention or graduation [...] otherwise, that's a financial loss to them". (P6)*

***Equity and access goals**: "The current model still has [...] the goal of equity [...] to really ensure that higher education is accessible". (P6)*

Conflating education goals with short-term metrics

- Education institutions can have competing goals, e.g. in admissions. These tensions tend not to be acknowledged in ML4Ed projects.

> ***Efficiency goal****: Universities "want students who would have higher rates of retention or graduation [...] otherwise, that's a financial loss to them". (P6)*

> ***Equity and access goals****: "The current model still has [...] the goal of equity [...] to really ensure that higher education is accessible". (P6)*

- Needs of key stakeholders could be left out.

Conflating education goals with short-term metrics

- Education institutions can have competing goals, e.g. in admissions. These tensions tend not to be acknowledged in ML4Ed projects.

> ***Efficiency goal****: Universities "want students who would have higher rates of retention or graduation […] otherwise, that's a financial loss to them". (P6)*

> ***Equity and access goals****: "The current model still has […] the goal of equity […] to really ensure that higher education is accessible". (P6)*

- Needs of key stakeholders could be left out.

> *P7: "Whose needs?* ***The student's needs, probably not.*** *[…] For the faculty, yeah, it's working well because what they want is to spend less time and get high quality students admitted."*

Limitations of a single prediction target

Limitations of a single prediction target

- Qualitative aspects of individual experience and narrative are irreducible to one or few quantitative targets. E.g. admissions, essay scoring.

Limitations of a single prediction target

- Qualitative aspects of individual experience and narrative are irreducible to one or few quantitative targets. E.g. admissions, essay scoring.

> *P3: "Holistic admission is not putting all these characteristics into a prediction model [to see] what's the probability [of admission]." Instead, the goal is to "**restore the actual life experience**, and [...] difficulties they overcome [...] to arrive where they are."*

Limitations of a single prediction target

- Qualitative aspects of individual experience and narrative are irreducible to one or few quantitative targets. E.g. admissions, essay scoring.

*P3: "Holistic admission is not putting all these characteristics into a prediction model [to see] what's the probability [of admission]." Instead, the goal is to "**restore the actual life experience**, and [...] difficulties they overcome [...] to arrive where they are."*

Limitations of a single prediction target

- Qualitative aspects of individual experience and narrative are irreducible to one or few quantitative targets. E.g. admissions, essay scoring.

> *P3: "Holistic admission is not putting all these characteristics into a prediction model [to see] what's the probability [of admission]." Instead, the goal is to "**restore the actual life experience**, and [...] difficulties they overcome [...] to arrive where they are."*

Limitations of a single prediction target

- Qualitative aspects of individual experience and narrative are irreducible to one or few quantitative targets. E.g. admissions, essay scoring.

*P3: "Holistic admission is not putting all these characteristics into a prediction model [to see] what's the probability [of admission]." Instead, the goal is to "**restore the actual life experience**, and [...] difficulties they overcome [...] to arrive where they are."*

*P3: Instead of giving **"an explicit ranking,"** the algorithmic system could "give summary information to the officers".*

Consequential choices in the prediction target

Consequential choices in the prediction target

- When predicting test outcomes, is the choice of test aligned with goal?

Consequential choices in the prediction target

- When predicting test outcomes, is the choice of test aligned with goal?

> *P4: "[A generic standardized test] more just measures [if]* ***they had more opportunity or more privilege in their background*** *than actually measuring if* ***they studied the material or if they learned.****"*

Consequential choices in the prediction target

- When predicting test outcomes, is the choice of test aligned with goal?

> *P4: "[A generic standardized test] more just measures [if]* ***they had more opportunity or more privilege in their background*** *than actually measuring if* ***they studied the material or if they learned."***

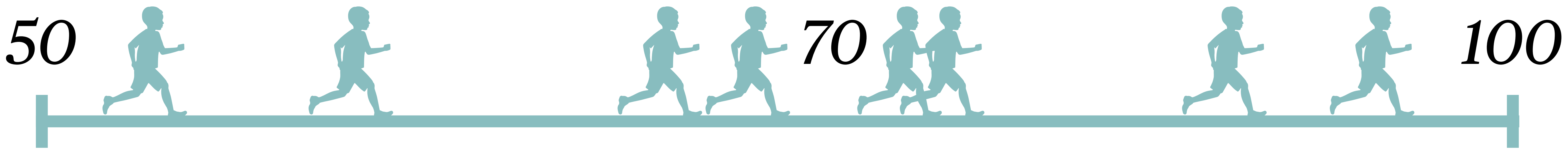
- Arbitrary thresholding to a binary target

Consequential choices in the prediction target

- When predicting test outcomes, is the choice of test aligned with goal?

> *P4: "[A generic standardized test] more just measures [if]* ***they had more opportunity or more privilege in their background*** *than actually measuring if* ***they studied the material or if they learned."***

- Arbitrary thresholding to a binary target

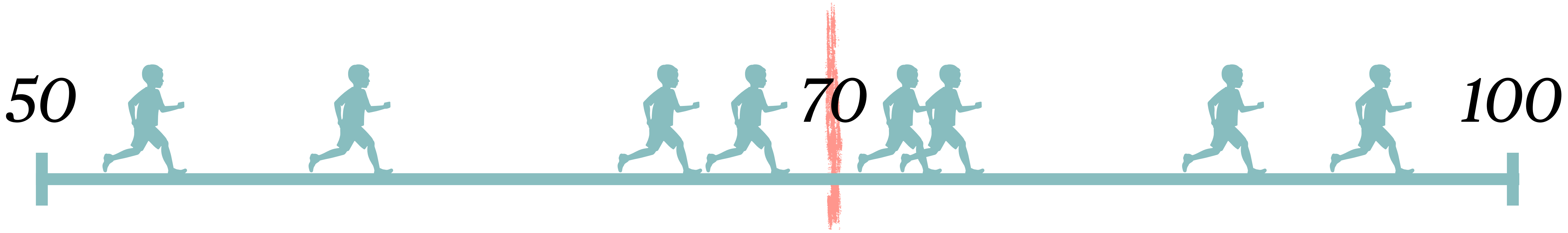


Consequential choices in the prediction target

- When predicting test outcomes, is the choice of test aligned with goal?

> *P4: "[A generic standardized test] more just measures [if]* ***they had more opportunity or more privilege in their background*** *than actually measuring if* ***they studied the material or if they learned."***

- Arbitrary thresholding to a binary target

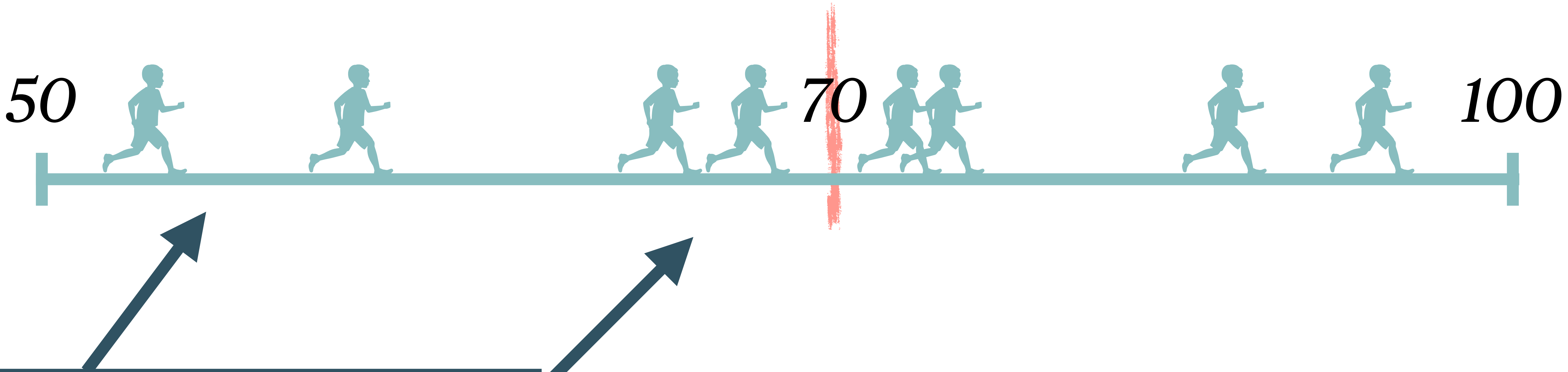


Consequential choices in the prediction target

- When predicting test outcomes, is the choice of test aligned with goal?

> *P4: "[A generic standardized test] more just measures [if]* ***they had more opportunity or more privilege in their background*** *than actually measuring if* ***they studied the material or if they learned."***

- Arbitrary thresholding to a binary target



> *P8: "those are different groups"*

Design inputs with education equity in mind

Design inputs with education equity in mind

- Participants emphasized the importance of acknowledging and addressing **structural inequality** in education-related research.

Design inputs with education equity in mind

- Participants emphasized the importance of acknowledging and addressing **structural inequality** in education-related research.
  - ML technologies that fail to consider structural inequality can actually widen gaps in education access (examples in paper)

Design inputs with education equity in mind

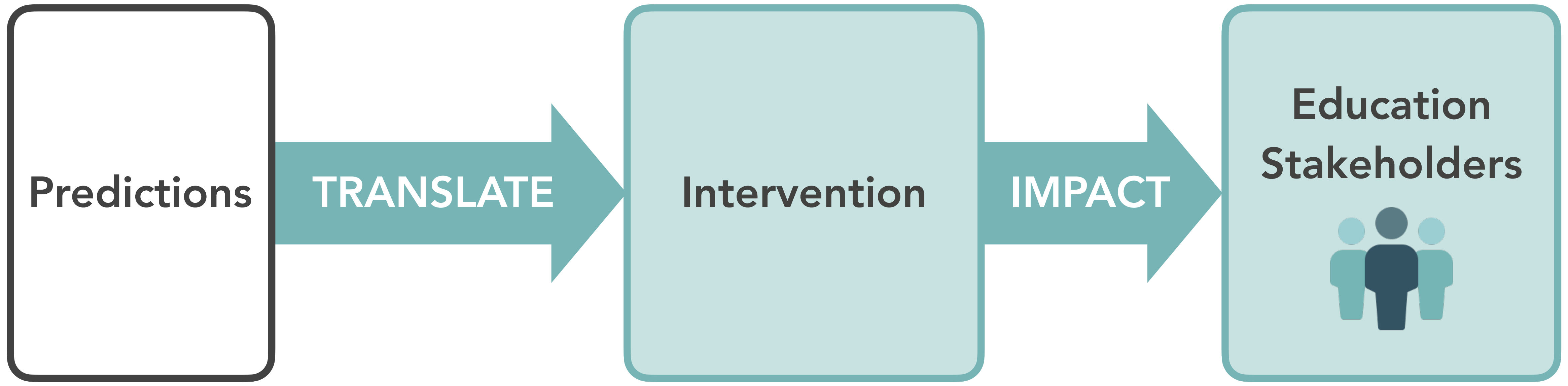
- Participants emphasized the importance of acknowledging and addressing **structural inequality** in education-related research.
    - ML technologies that fail to consider structural inequality can actually widen gaps in education access (examples in paper)
- **Predicting with race and ethnicity attributes** without accounting for the ways in which resources and opportunity are allocated by race and ethnicity can lead to erroneous conclusions about the role of race and ethnicity in educational outcomes.

Design inputs with education equity in mind

- Participants emphasized the importance of acknowledging and addressing **structural inequality** in education-related research.
    - ML technologies that fail to consider structural inequality can actually widen gaps in education access (examples in paper)
- **Predicting with race and ethnicity attributes** without accounting for the ways in which resources and opportunity are allocated by race and ethnicity can lead to erroneous conclusions about the role of race and ethnicity in educational outcomes.

> *P04: "**[Race] was** a little **too nuanced** [...] But a researcher would never think of it that way, right? They [...] **want to get the best prediction possible**".*

Part 2: Translating Predictions to Interventions



Better Predictions ≠ Better Interventions

Better Predictions ≠ Better Interventions

- Most machine learning papers tout improvements in prediction accuracy, e.g. predicting student dropout risk, without considering interventions.

Better Predictions ≠ Better Interventions

- Most machine learning papers tout improvements in prediction accuracy, e.g. predicting student dropout risk, without considering interventions.

- Participants question the value of predicting “better”:

Better Predictions ≠ Better Interventions

- Most machine learning papers tout improvements in prediction accuracy, e.g. predicting student dropout risk, without considering interventions.
- Participants question the value of predicting “better”:

> *P4: Computer science papers focus on asking “‘can we predict 62% instead of 61%,’ but* ***nobody had anything to do that would actually help*** *these people.”*

Better Predictions ≠ Better Interventions

- Most machine learning papers tout improvements in prediction accuracy, e.g. predicting student dropout risk, without considering interventions.
- Participants question the value of predicting “better”:

P4: Computer science papers focus on asking “‘can we predict 62% instead of 61%,’ but ***nobody had anything to do that would actually help*** *these people.”*

*P5: “****You don't improve things by predicting them better.****”*

Better Predictions ≠ Better Interventions

- Most machine learning papers tout improvements in prediction accuracy, e.g. predicting student dropout risk, without considering interventions.
- Participants question the value of predicting “better”:

P4: Computer science papers focus on asking “‘can we predict 62% instead of 61%,’ but ***nobody had anything to do that would actually help*** *these people.”*

*P5: “****You don't improve things by predicting them better.****”*

P6: “Even if you tell them [...] that they have a 97% chance of dropping out based on our training data, ***that’s a difficult thing to take in*** *especially in the public schools [...] [where it’s] very difficult to find good teachers for those students.”*

Harms of naive interventions

Harms of naive interventions

Risk prediction:

- Showing failure risk predictions to students can de-motivate students and lead to a self fulfilling prophecy.

Harms of naive interventions

Risk prediction:

- Showing failure risk predictions to students can de-motivate students and lead to a self fulfilling prophecy.

> *P4: "It doesn't seem well connected to research about **what motivates people.**"*

Harms of naive interventions

Risk prediction:

- Showing failure risk predictions to students can de-motivate students and lead to a self fulfilling prophecy.

*P4: “It doesn't seem well connected to research about **what motivates people.**”*

*P1: “There is a tendency for a lot of these systems to **stigmatize students**.”*

Harms of naive interventions

Risk prediction:

- Showing failure risk predictions to students can de-motivate students and lead to a self fulfilling prophecy.

> *P4: "It doesn't seem well connected to research about **what motivates people.**"*

> *P1: "There is a tendency for a lot of these systems to **stigmatize students**."*

- Scores affect behavior of teachers, family, and institutions as well.

Harms of naive interventions

Risk prediction:

- Showing failure risk predictions to students can de-motivate students and lead to a self fulfilling prophecy.

> *P4: "It doesn't seem well connected to research about **what motivates people.**"*

> *P1: "There is a tendency for a lot of these systems to **stigmatize students**."*

- Scores affect behavior of teachers, family, and institutions as well.

> *P6: A teacher "might allocate more of their limited time to other students rather than a student that the model seems to predict that they will not graduate."*

Towards intervention-aware prediction

Towards intervention-aware prediction

Actionability vs. Interpretability

- Participants repeatedly pointed to the value of building models that provide actionable insights.

Towards intervention-aware prediction

Actionability vs. Interpretability

- Participants repeatedly pointed to the value of building models that provide actionable insights.

P9: "We don't have to say, 'You're not going to succeed.' We can say, 'Let's talk about what are the decisions that you need to make, ***what is the pathway forward that will allow you to succeed.'"***

Towards intervention-aware prediction

Actionability vs. Interpretability

- Participants repeatedly pointed to the value of building models that provide actionable insights.

P9: "We don't have to say, 'You're not going to succeed.' We can say, 'Let's talk about what are the decisions that you need to make, ***what is the pathway forward that will allow you to succeed.'"***

*P10: "'You're in the 10th percentile for something' sounds different than '****we're worried because you've been absent a lot.'"***

Towards intervention-aware prediction

Towards intervention-aware prediction

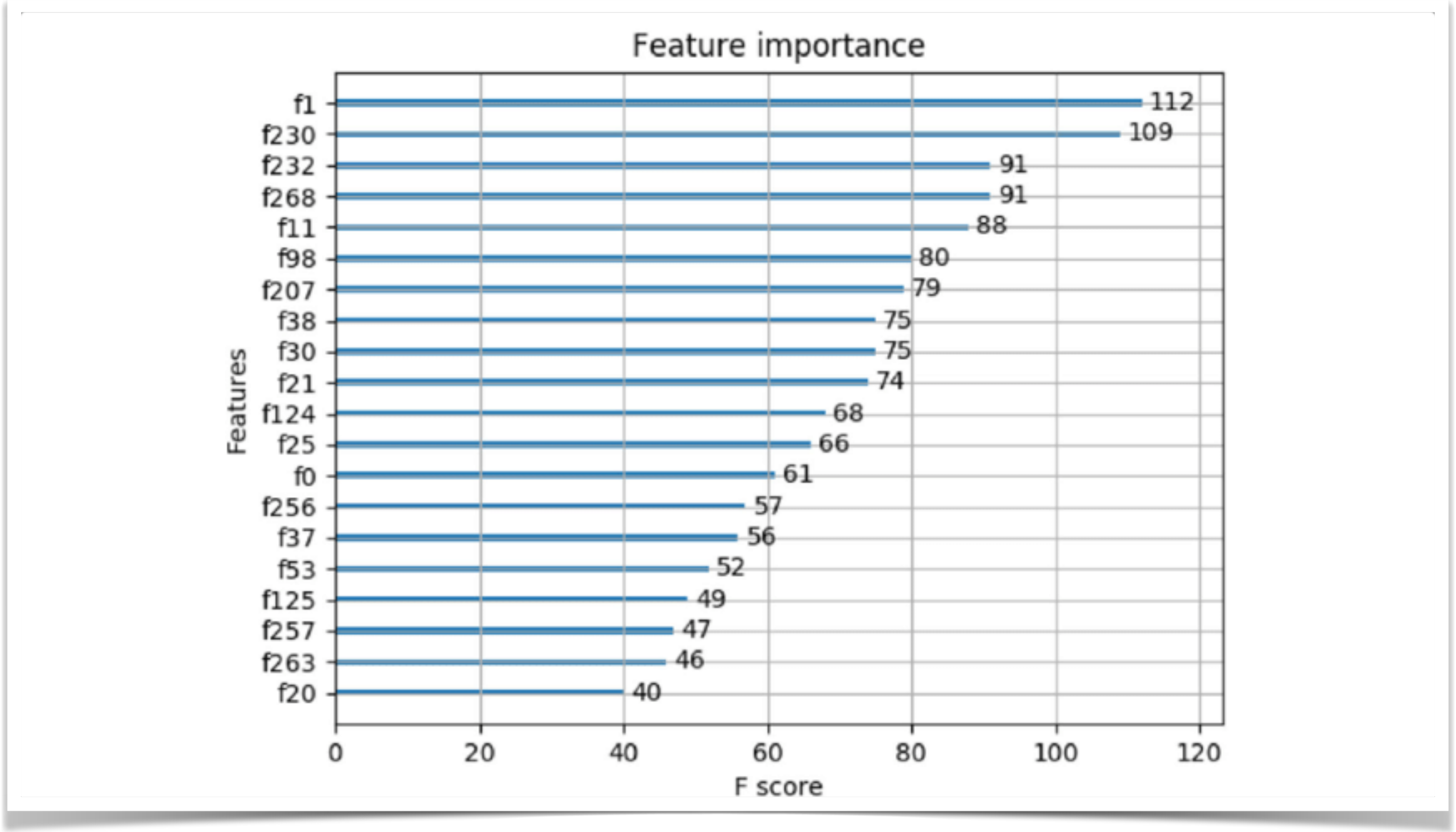
Actionability vs. Interpretability

- Educators are less interested in “interpretability” in terms of feature importance.

Towards intervention-aware prediction

<u>Actionability vs. Interpretability</u>

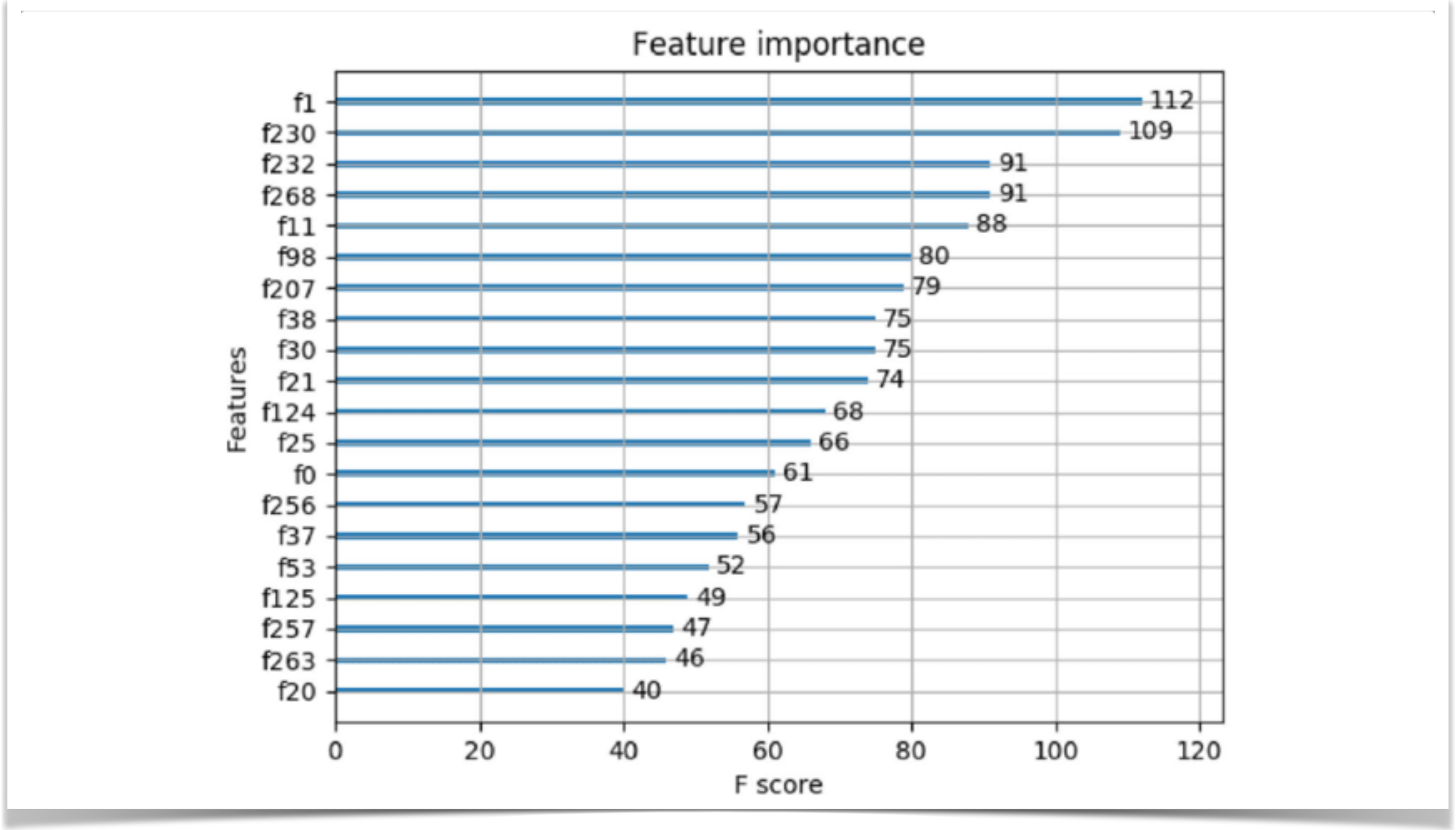
- Educators are less interested in “interpretability” in terms of feature importance.



Towards intervention-aware prediction

<u>Actionability vs. Interpretability</u>

- Educators are less interested in "interpretability" in terms of feature importance.



*P11: "What is most interesting about to me is not, 'I wonder if the demographic factors matter more than the behavioral factors.' To me it's more about, '**what can we actually do to help kids get off the trajectory they're on if they're not on a good trajectory.**'"*

Towards intervention-aware prediction

Towards intervention-aware prediction

Design to empower human operators

Towards intervention-aware prediction

Design to empower human operators

- Risk prediction systems could add value if they were designed with the teacher's partnership in mind.

Towards intervention-aware prediction

Design to empower human operators

- Risk prediction systems could add value if they were designed with the teacher's partnership in mind.

> *P04:* ***When the teachers have the power and resources to intervene,*** *risk prediction algorithms can "help [teachers] catch some kids that maybe had fallen below the radar before and giving them another source of data on that".*

Towards intervention-aware prediction

<u>Design to empower human operators</u>

- Risk prediction systems could add value if they were designed with the teacher's partnership in mind.

> *P04: **When the teachers have the power and resources to intervene**, risk prediction algorithms can "help [teachers] catch some kids that maybe had fallen below the radar before and giving them another source of data on that".*

- The goal of predicting students that the teachers would have overlooked is **different from the standard goal** of achieving high predictive accuracy for the entire student population

Towards intervention-aware prediction

Towards intervention-aware prediction

Design to empower human operators

Towards intervention-aware prediction

<u>Design to empower human operators</u>

- Effective usage of automated chatbots requires knowing what to ask.

Towards intervention-aware prediction

Design to empower human operators

- Effective usage of automated chatbots requires knowing what to ask.
- Advisors “fill in the gaps for students” to help them “envision a pathway” from a relatively vague conversation.

Towards intervention-aware prediction

<u>Design to empower human operators</u>

- Effective usage of automated chatbots requires knowing what to ask.
- Advisors “fill in the gaps for students” to help them “envision a pathway” from a relatively vague conversation.

Towards intervention-aware prediction

Design to empower human operators

- Effective usage of automated chatbots requires knowing what to ask.
- Advisors “fill in the gaps for students” to help them “envision a pathway” from a relatively vague conversation.

P9: “You're not going to be a nurse, sorry, like, good luck”

Towards intervention-aware prediction

<u>Design to empower human operators</u>

- Effective usage of automated chatbots requires knowing what to ask.
- Advisors “fill in the gaps for students” to help them “envision a pathway” from a relatively vague conversation.

P9: “We have respiratory therapy, or we have nutrition, ...all these other healthcare disciplines that might allow you to help people, to work in healthcare, and to get a job, which [are] [...] three boxes you said you wanted to check.”

P9: “You're not going to be a nurse, sorry, like, good luck”

Towards intervention-aware prediction

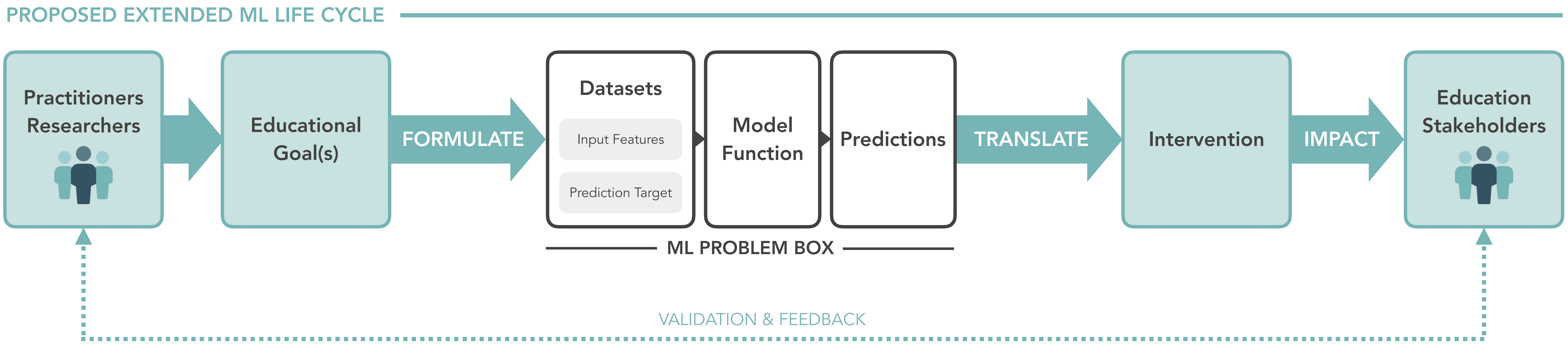
Design to empower human operators

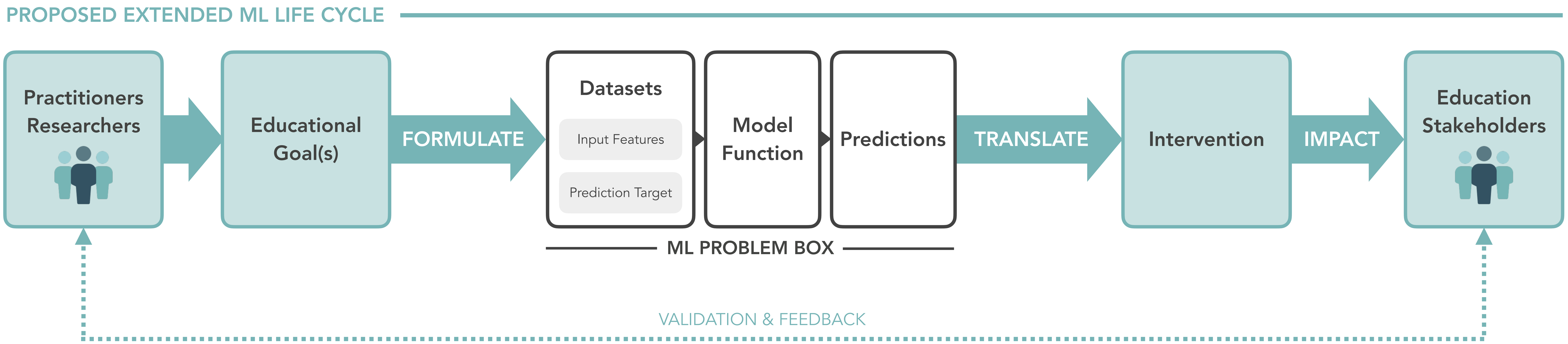
- Effective usage of automated chatbots requires knowing what to ask.
- Advisors “fill in the gaps for students” to help them “envision a pathway relatively vague conversation.

P9: “That sounds great. I had no idea what an occupational therapist even was.”

P9: “We have respiratory therapy, or we have nutrition, ...all these other healthcare disciplines that might allow you to help people, to work in healthcare, and to get a job, which [are] [...] three boxes you said you wanted to check.”

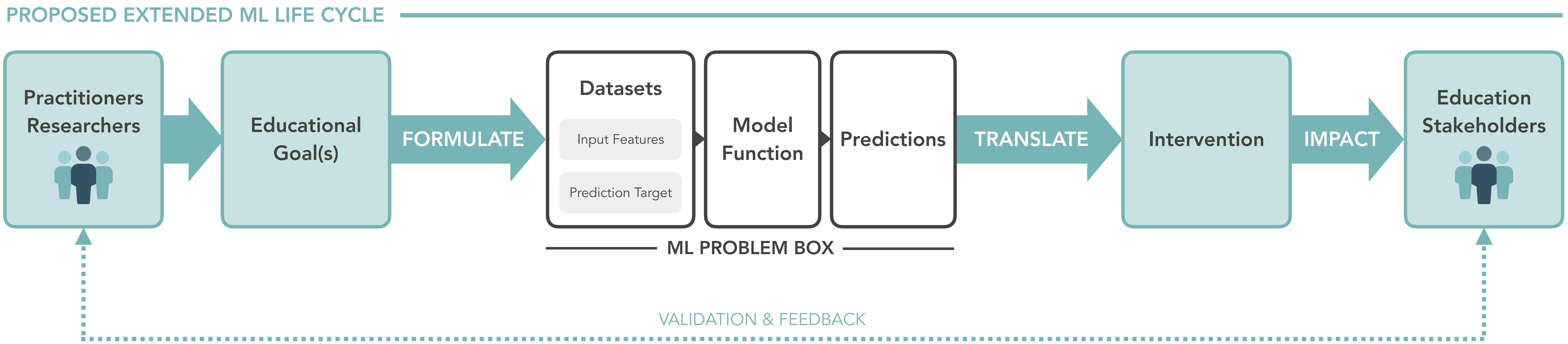
P9: “You're not going to be a nurse, sorry, like, good luck”


PROPOSED EXTENDED ML LIFE CYCLE
Practitioners
Researchers
Educational
Goal(s)
FORMULATE
Datasets
Input Features
Prediction Target
Model
Function
Predictions
ML PROBLEM BOX
TRANSLATE
Intervention
IMPACT
Education
Stakeholders
VALIDATION & FEEDBACK



Part 1: Translating Education Goals via Problem Formulation

- Conflating education goals with short-term metrics
- Limitations of a single prediction target
- Design inputs with education equity in mind



Part 1: Translating Education Goals via Problem Formulation

- Conflating education goals with short-term metrics
- Limitations of a single prediction target
- Design inputs with education equity in mind

Part 2: Translating Predictions to Interventions

- Gap between predictions and interventions
- Harms of naive interventions from prediction
- Towards intervention-aware prediction

Future work

Future work

- Gaps in (1) ML problem formulation and (2) from predictions to interventions

Future work

- Gaps in (1) ML problem formulation and (2) from predictions to interventions
  - *Beyond single objectives:* How to navigate the ecosystem of multiple education goals and stakeholders in the formulation of ML tasks?

Future work

- Gaps in (1) ML problem formulation and (2) from predictions to interventions
  - *Beyond single objectives:* How to navigate the ecosystem of multiple education goals and stakeholders in the formulation of ML tasks?
  - *Beyond prediction*: How to produce actionable insights with ML tasks?

Future work

- Gaps in (1) ML problem formulation and (2) from predictions to interventions
  - *Beyond single objectives:* How to navigate the ecosystem of multiple education goals and stakeholders in the formulation of ML tasks?
  - *Beyond prediction*: How to produce actionable insights with ML tasks?
- Extensions of our methodology, e.g. fieldwork and focus groups with direct stakeholders (students and teachers), quantitative statistics from ML4Ed papers.

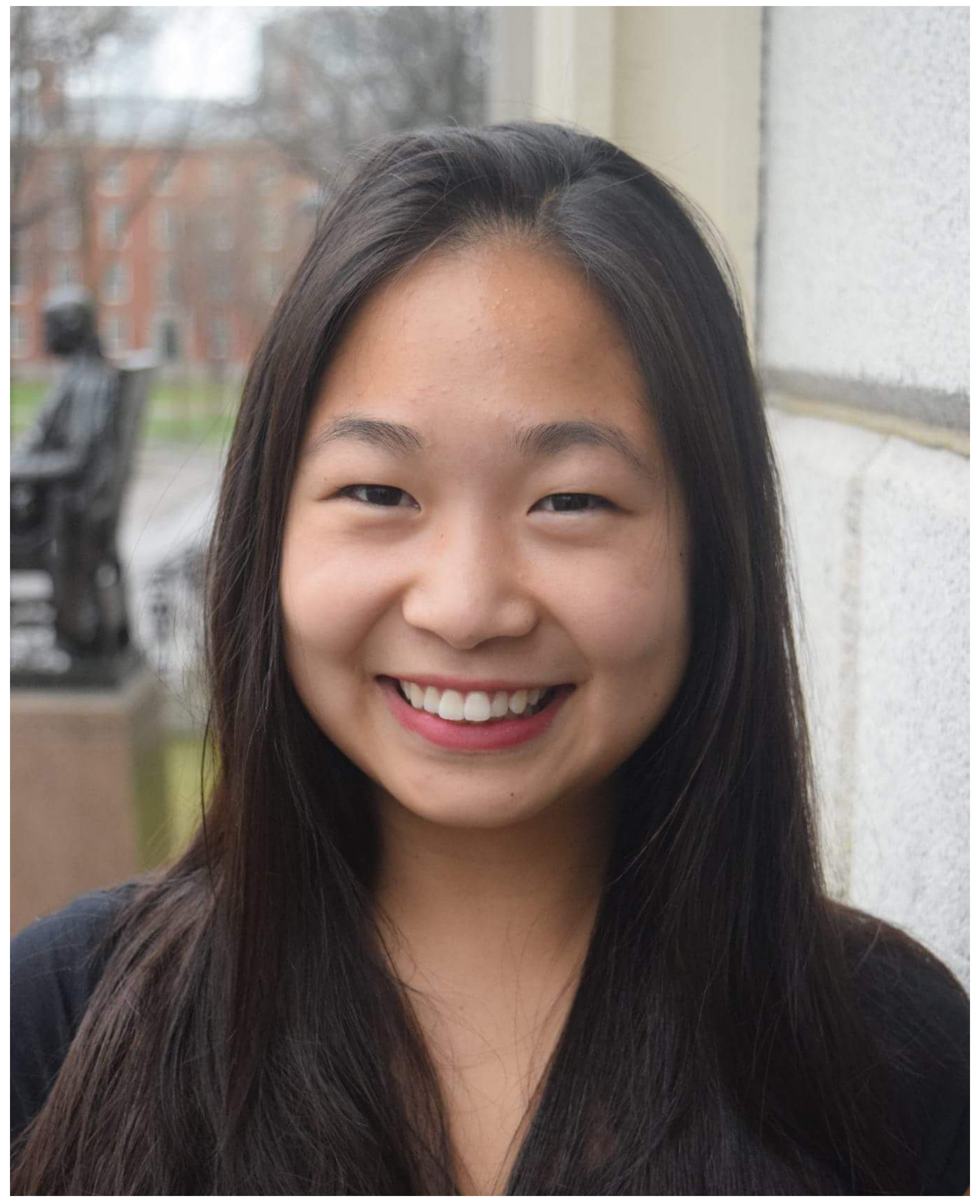
Future work

- Gaps in (1) ML problem formulation and (2) from predictions to interventions
  - *Beyond single objectives:* How to navigate the ecosystem of multiple education goals and stakeholders in the formulation of ML tasks?
  - *Beyond prediction*: How to produce actionable insights with ML tasks?
- Extensions of our methodology, e.g. fieldwork and focus groups with direct stakeholders (students and teachers), quantitative statistics from ML4Ed papers.
- *Beyond education:* Using this methodology to examine other domains of ML4SG

Future work

- Gaps in (1) ML problem formulation and (2) from predictions to interventions
  - *Beyond single objectives:* How to navigate the ecosystem of multiple education goals and stakeholders in the formulation of ML tasks?
  - *Beyond prediction*: How to produce actionable insights with ML tasks?
- Extensions of our methodology, e.g. fieldwork and focus groups with direct stakeholders (students and teachers), quantitative statistics from ML4Ed papers.
- *Beyond education:* Using this methodology to examine other domains of ML4SG
  - Healthcare, Criminal justice/legal system, Social services sector, Environmental protection

Thank you!



Lydia T. Liu*
UC Berkeley EECS

Serena Wang*
UC Berkeley EECS

Rediet Abebe†
UC Berkeley EECS

Tolani Britton†
UC Berkeley GSE

lydiatliu.com | lydiatliu (at) berkeley (dot) edu

* Equal authorship †Equal advisorship