



"Af hverju notið þið ekki bara ChatGPT?"

Þegar gervigreindin er lækningatæki

Ský, 14 October 2026

Jón Skírnir Ágústsson, PhD

VP AI & Data Research

AI at Nox since 2015

2015

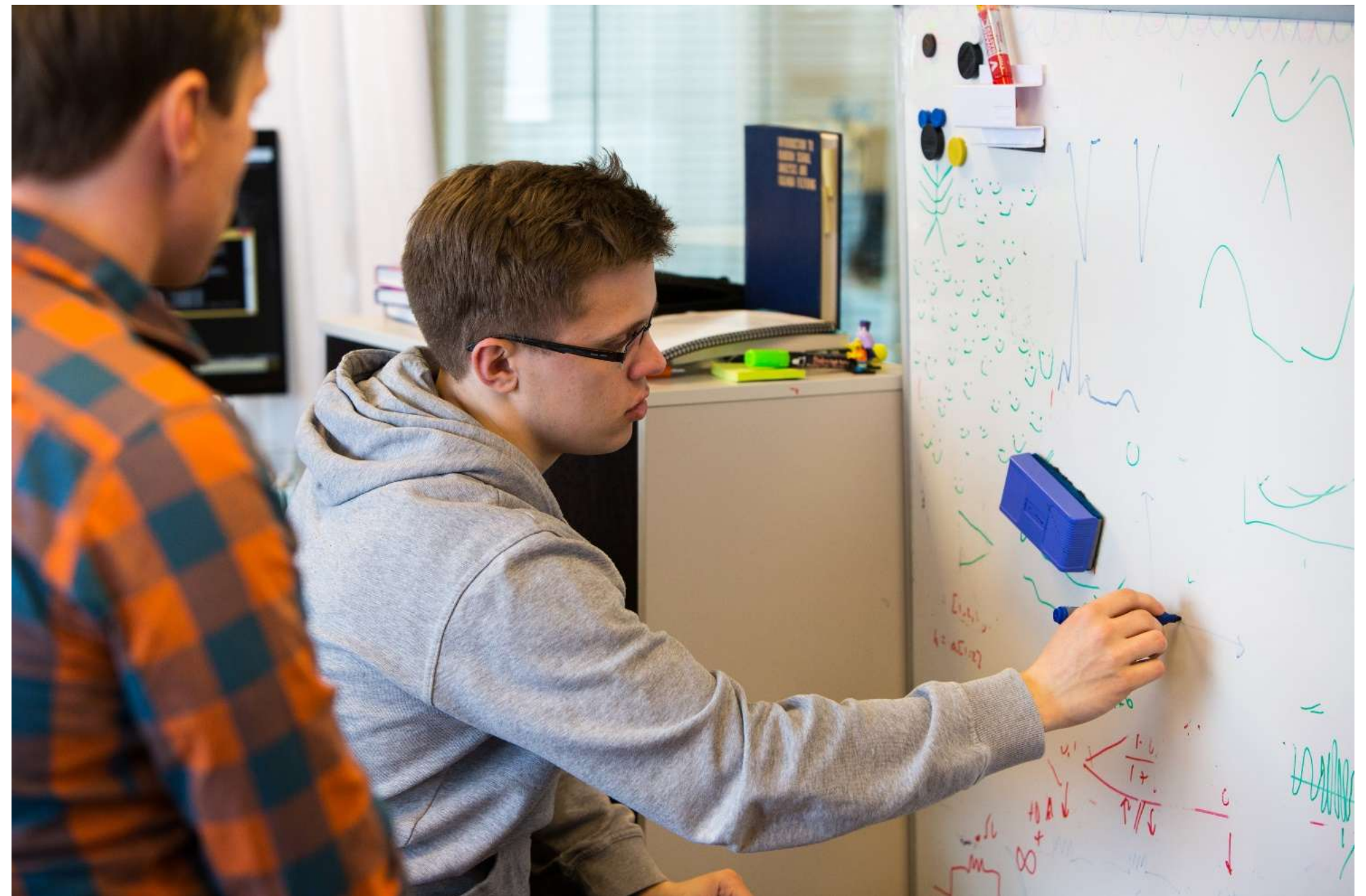
First deep neural network

2016–2021

Sleep staging, BodySleep and arousal detection in Noxturnal

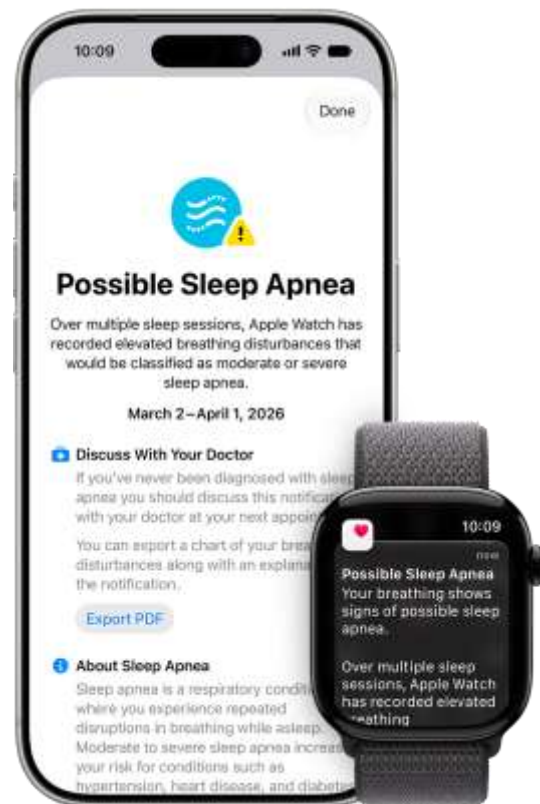
2025

Two FDA 510(k) clearances for DeepRESP



Different intended uses, different requirements

Intended medical use determines requirements, risks and performance criteria



Apple

A medical device within a consumer product

Possible sleep apnea in adults without a prior diagnosis

Repeated measurements in everyday use

No notification does not rule out apnea

Fewer false notifications
High specificity, lower sensitivity



Nox

Clinical sleep apnea assessment in patients referred for testing

Results can lead to long-term treatment

Missed cases can delay treatment

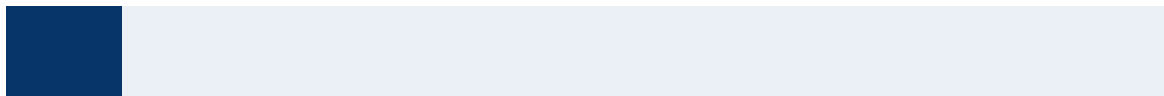
Diagnosis and treatment
Balance sensitivity and specificity

Different clinical priorities, different trade-offs

Device A limits false alerts. Device B balances finding cases with false positives.

Device A: notification

10% have the condition



90% do not have the condition

Device B: diagnosis

60% have the condition



40% do not have the condition

Accuracy	95.1%	90%
----------	-------	-----

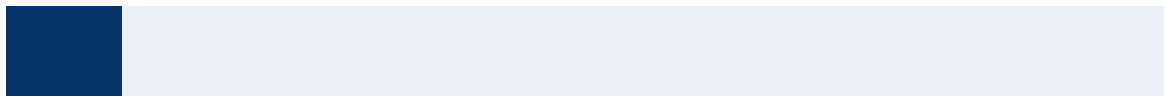
Hypothetical examples inspired by notification and diagnostic priorities. Not measured Apple or Nox results.

Different clinical priorities, different trade-offs

Device A limits false alerts. Device B balances finding cases with false positives.

Device A: notification

10% have the condition



90% do not have the condition

Device B: diagnosis

60% have the condition



40% do not have the condition

Accuracy	95.1%	90%
Sensitivity	60%	90%
Specificity	99%	90%

Sensitivity: identifies people who have the condition

Specificity: correctly clears people who do not have the condition

Hypothetical examples inspired by notification and diagnostic priorities. Not measured Apple or Nox results.

Documentation behind an AI medical device

DeepRESP v1.0 and v2.0: figures from the Nox quality management system

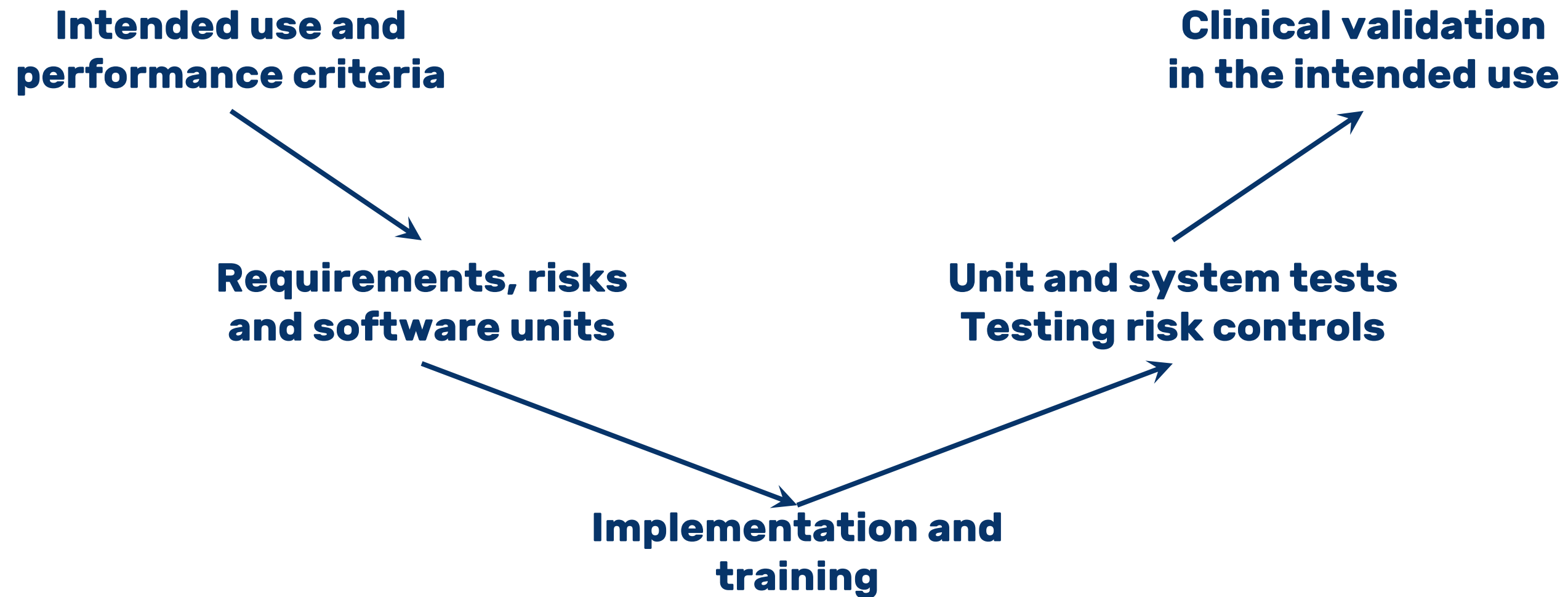
	DeepRESP v1.0	DeepRESP v2.0
Python lines of code	101,977	119,538
Design documents	60	42
Documentation pages	≈1,650	≈1,460
Estimated lines of text	≈66,000	≈58,000

Requirements, software design, risk management and cybersecurity
Verification, clinical validation and project approval

Source: MasterControl and project DHF registers. Text-line estimates assume 40 lines/page.

DeepRESP: requirements, development and testing

Clinical purpose and performance criteria guide the development process



Spec-driven development: define success before final validation

Sleep studies: overnight physiological recordings



Home sleep apnea test (HSAT)

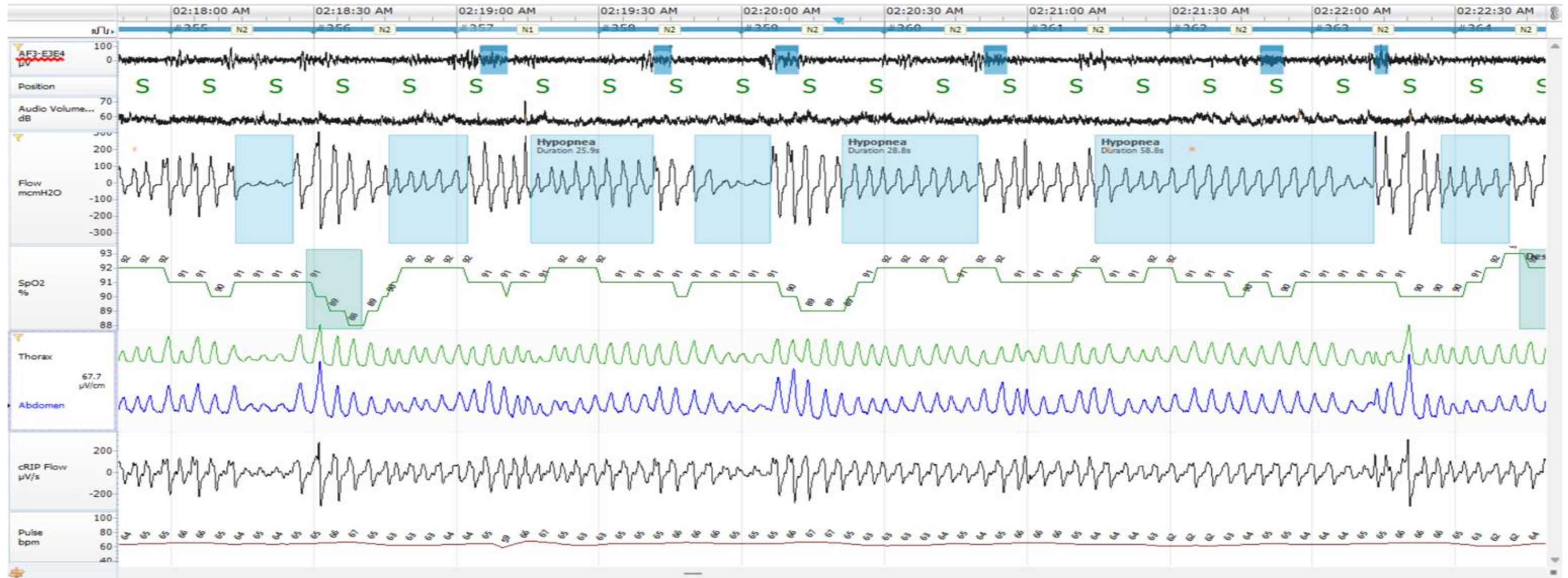
Nasal airflow
Chest and abdominal breathing effort
Oxygen saturation
Movement and sound



Polysomnography (PSG)

Additional measurements:
Brain activity (EEG)
Muscle activity and leg movements
Heart activity (ECG)

Sleep studies: signals and clinical interpretation



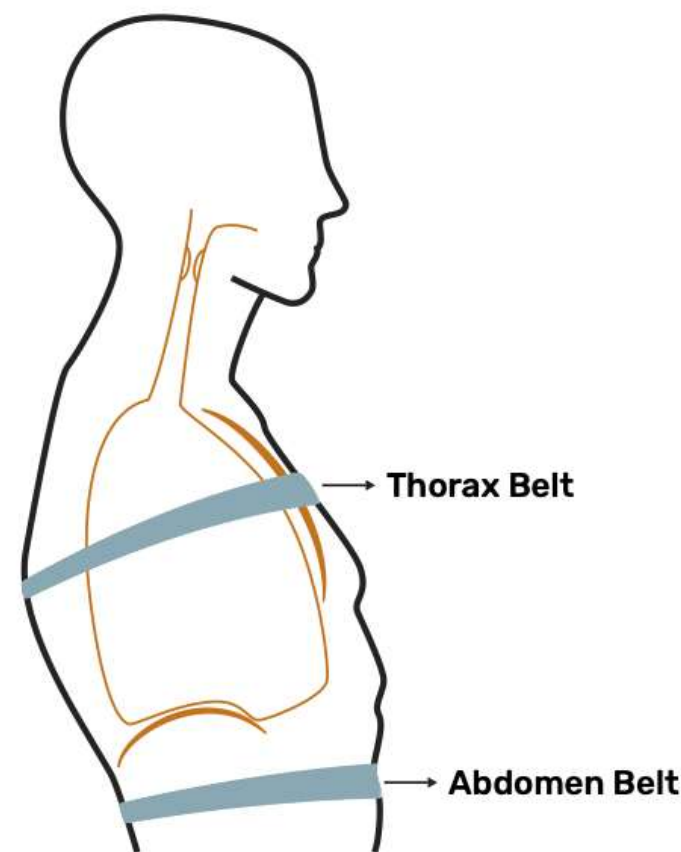
Reference: PSG scoring with EEG

**Clinical workflow:
review by a healthcare professional**

nox medical

How does breathing reflect sleep state?

We used our experience in EEG scoring to develop the BodySleep neural network



Wake

Variable breathing and voluntary control

NREM

More regular breathing, driven by metabolic demand

REM

Variable breathing and reduced muscle tone

Arousal

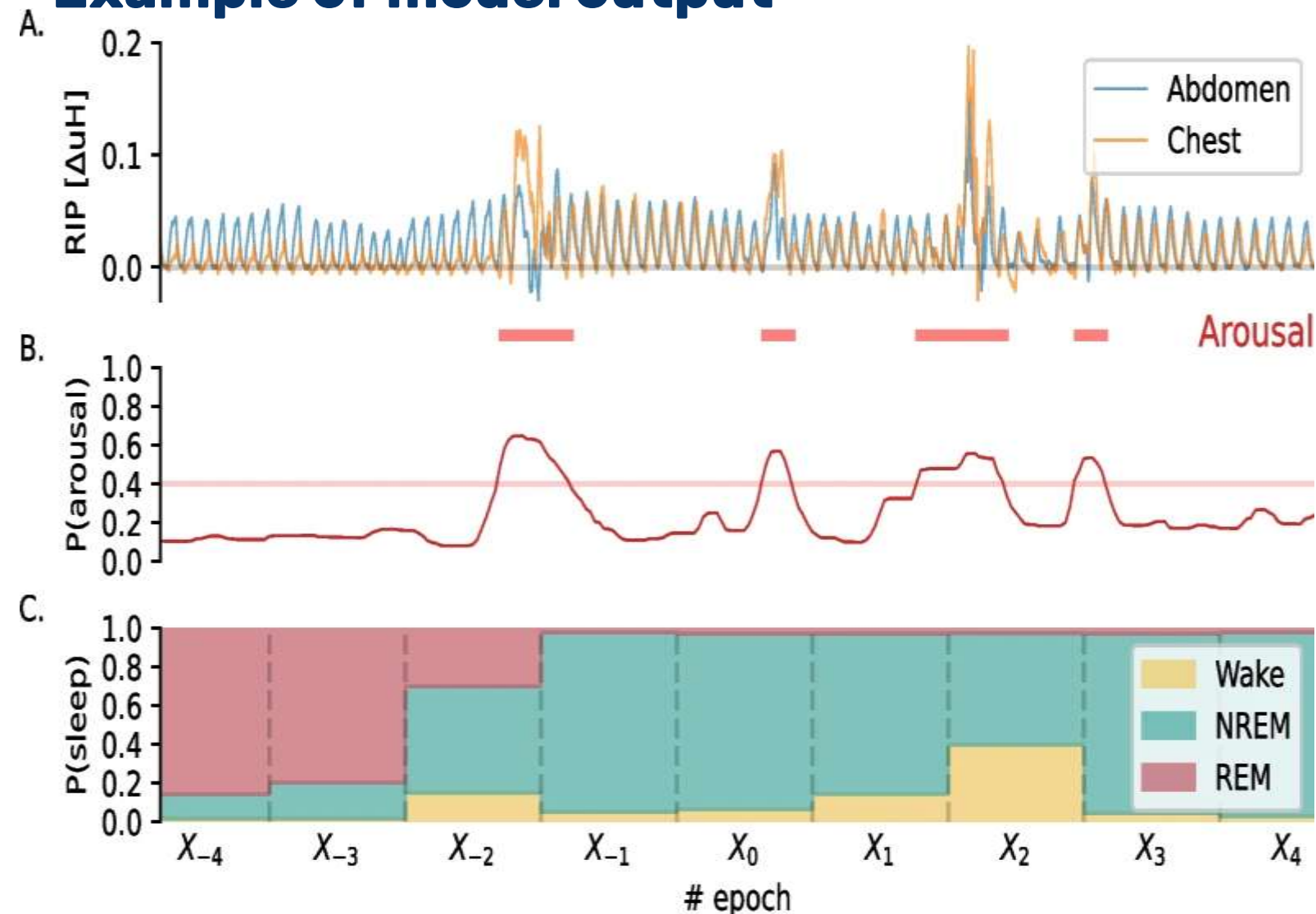
Changes in breathing control and increased ventilation

RIP belts measure chest and abdominal movement. These physiological changes support sleep assessment from breathing.

BodySleep: sleep and arousals from breathing

Inputs: RIP signals. Reference: PSG scoring with EEG.

Example of model output



Finnsso et al. (2025), Fig. 2.

An output example, not a performance evaluation.

Human review of AI output

- Familiar physiological signals
- Familiar sleep and arousal labels
- Model probabilities help reviewers assess the output

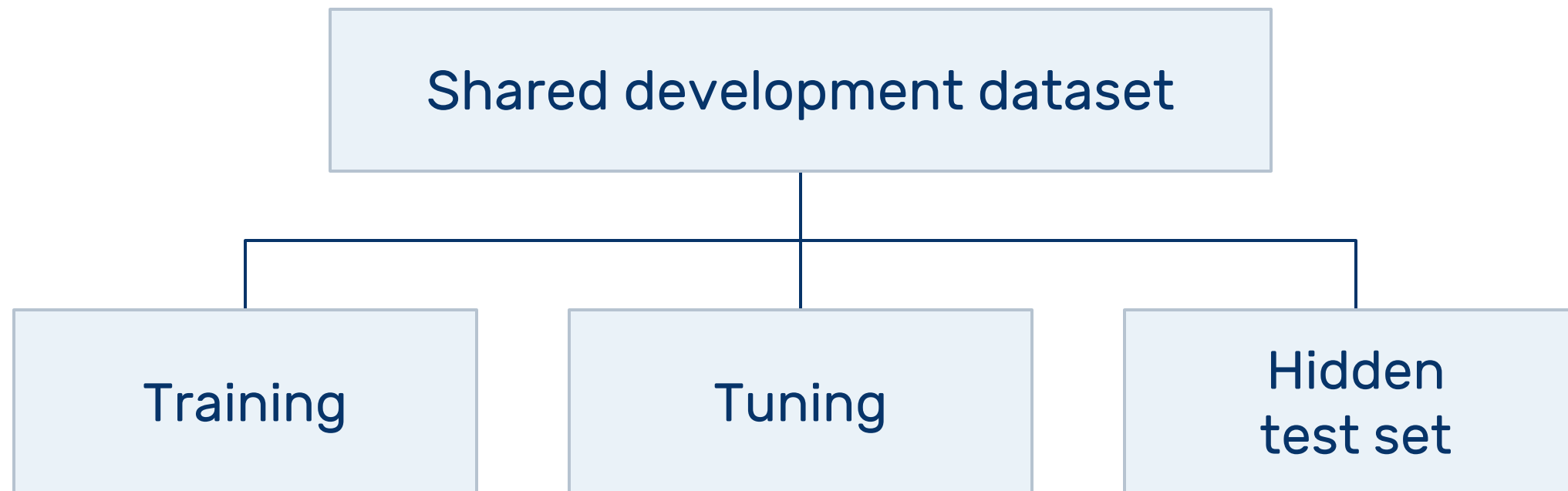
IMDRF GMLP, principle 7: performance of the human-AI team

Transparent output supports clinical review

Independent data for clinical validation

Final validation uses independent data from the intended patient population

Development data



Unseen patients can still share hospitals and recording conditions

Data independence and population coverage are separate requirements

Clinical validation



No data from these hospitals in the development dataset

Different body types, ages and clinical conditions

IMDRF GMLP (2025), principles 3, 4 and 8. Illustration of the Nox approach.

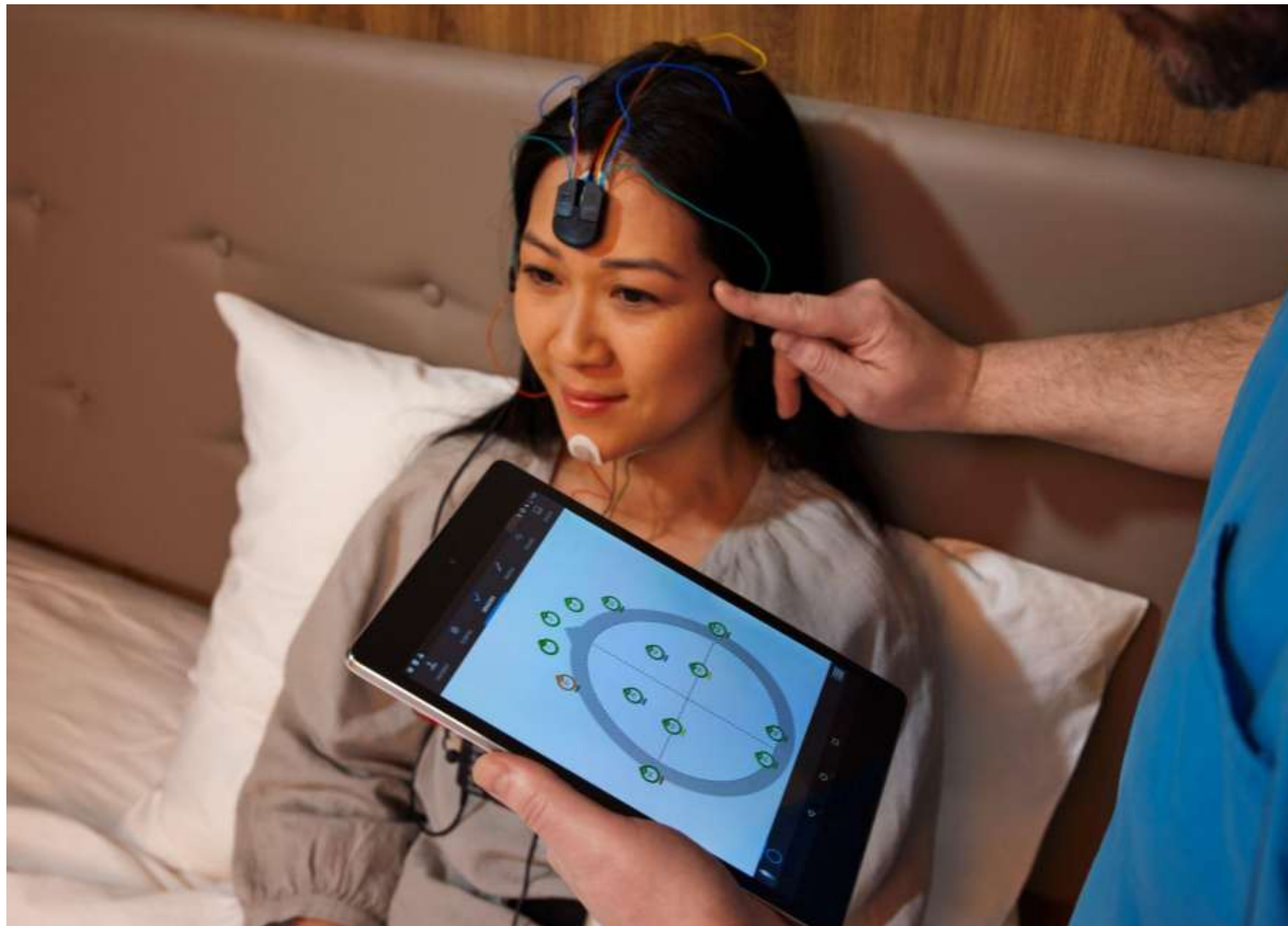
How we use AI in medical device development

Examples from daily development work at Nox

Standards and documents	Code and review	V&V and clinical evidence
Read standards and guidance and assess changes	Assist coding under the development team's control	Design verification activities and clinical validation
Check procedures and documents against applicable requirements	Review pull requests for errors and coding style	Review clinical protocols and reports against regulatory expectations
Write and maintain documentation	Speed up UI testing	

Traceability: connect requirements to implementation, tests and risk controls

Why did we develop our own models?



A specialized task

Physiological time series
Sleep stages and timed events
In 2015, language and image approaches
did not fully fit our task

Understanding and control

Training data and edge cases
Known performance for each release
Control of the model and input signals

Third-party models need the same
supporting evidence and change control
as our own.



Jón Skírnir Ágústsson
Ský, 14 October 2026