**Engifi ↔ O\*NET Mapping: Technical Specification (v1)**

**0) Scope**

Map public O\*NET occupation data into Engifi’s feature space to set **role-level mean profiles** (z-means) for inputs used in synthetic data generation and later modeling.

**1) Input Sources (O\*NET data domains)**

* **Abilities**: Inductive Reasoning, Deductive Reasoning, Problem Sensitivity, Information Ordering, Perceptual Speed, Category Flexibility, Written/Oral Comprehension, etc.
* **Work Styles**: Dependability, Cooperation, Attention to Detail, Integrity, Stress Tolerance, Persistence, Adaptability/Flexibility.
* **Interests**: RIASEC—Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), Conventional (C).
* **Work Context / Work Activities** (for autonomy/structure proxy): Freedom to Make Decisions, Structured vs. Unstructured Work, Pace Determined by Speed of Equipment, Responsibility for Outcomes, Contact With Others, etc.

Note: Use the **“Importance”** or **“Level”** ratings (0–100) as provided by O\*NET for each occupation-scale pair. If both exist, prefer **Importance**; use **Level** as a sensitivity check.

**2) Output Targets (Engifi constructs)**

* **Cognitive**: Gf (Fluid), Gc (Crystallized), WM (Working Memory/Attention Control), Speed (Processing/Perceptual Speed), plus derived GMA (post-hoc PCA of [Gf,Gc,WM,Speed], not mapped directly).
* **Non-cognitive**: Conscientiousness (work ethic/self-regulation), Agreeableness (prosociality/cooperation), AutonomyPref (preference for autonomy vs. structure).
* **Interests**: R,I,A,S,E,C (RIASEC).
* **Values** (optional in v1): org-level vector; not directly from O\*NET (can be set separately).

**3) Deterministic Mapping Rules**

**3.1 Abilities → Cognitive**

|  |  |  |  |
| --- | --- | --- | --- |
| **Engifi** | **O\*NET Abilities (primary drivers)** | **O\*NET Abilities (secondary contributors)** | **Transform** |
| **Gf** | Inductive Reasoning, Deductive Reasoning | Category Flexibility, Problem Sensitivity | z from weighted mean (primary 0.7, secondary 0.3) |
| **Gc** | Written Comprehension, Oral Comprehension | Written Expression, Oral Expression | z from weighted mean (0.7 / 0.3) |
| **WM** | Memorization, Selective Attention | Information Ordering | z from weighted mean (0.7 / 0.3) |
| **Speed** | Perceptual Speed | Speed of Closure, Flexibility of Closure | z from weighted mean (0.7 / 0.3) |

**Weights (w):**

* Primary O\*NET scales weight **w₁=0.70** each (normalized if >1 scale).
* Secondary O\*NET scales weight **w₂=0.30** (normalized).
* If a secondary scale is missing, renormalize remaining weights to sum to 1.

**Computation per occupation o:**

score\_raw(Engifi\_k, o) = Σ\_i w\_i \* O\*NET\_score\_i(o)

### 3.2 Work Styles → Non-cognitive

|  |  |  |
| --- | --- | --- |
| **Engifi** | **O\*NET Work Styles** | **Rule** |
| **Conscientiousness** | Dependability, Attention to Detail, Persistence, Self-Control | Weighted mean: Dependability 0.5, Attention 0.2, Persistence 0.2, Self-Control 0.1 |
| **Agreeableness** | Cooperation, Concern for Others | Weighted mean: Cooperation 0.7, Concern 0.3 |

### 3.3 Work Context → AutonomyPref

* **Positive contributors to autonomy**: Freedom to Make Decisions (+), Structured vs. Unstructured Work (invert; more unstructured → more autonomy), Impact of Decisions.
* **Negative contributors (structure)**: Pace Determined by Equipment (+structure), Degree of Automation (+structure).

AutonomyPref\_raw =

+ 0.45 \* z(Freedom\_to\_Make\_Decisions)

+ 0.35 \* z(Unstructured\_Work) // = - z(Structured\_Work) if only “Structured” is available

+ 0.20 \* z(Impact\_of\_Decisions)

- 0.20 \* z(Pace\_by\_Equipment)

- 0.10 \* z(Degree\_of\_Automation)

### 3.4 Interests → RIASEC

Direct pass-through (after z-normalization):

R,I,A,S,E,C = z( RIASEC\_importance )

If only Holland **codes** (e.g., “IRC”) are available, assign heuristic means:

* Primary letter: +0.6, secondary: +0.3, tertiary: +0.1, others: −0.1 (then re-center to mean 0).

## 4) Normalization & Scaling

### 4.1 Z-normalization

For any O\*NET scale X with values on [0,100]:

z\_X(o) = ( X(o) - μ\_X ) / σ\_X

Where μ\_X and σ\_X are computed across the **full O\*NET occupation universe** for that scale. If unavailable, use proxy μ=50, σ=15, then replace with true μ/σ when pulled.

### 4.2 Composite z for each Engifi construct

If score\_raw is a weighted sum of multiple O\*NET scales:

1. Convert each input O\*NET scale to **z** first.
2. Compute a weighted sum of **z** values:

z\_comp(o) = Σ\_i w\_i \* z\_Xi(o)

1. Optionally **cap** extreme z to ±1.0 for realism in priors (avoid overly peaky means).

### 4.3 Rescaling to Engifi “H/M/L” shorthand

When needed for documentation only:

* **H = +0.6**, **M = 0.0**, **L = −0.6**.

## 5) Missing Data & Edge Handling

* **Missing O\*NET scale**: drop the term and **renormalize** weights in the composite.
* **Inconsistent scales (Importance vs Level)**: prefer **Importance**; if Level exists and diverges >1.0 z, average (Importance 0.7, Level 0.3) and flag for SME review.
* **Outliers**: cap final z at ±1.0 in the **prior** (uncapped later if you move to numeric O\*NET universe stats).

## 6) SME Adjustment (optional but recommended)

* Collect **pairwise AHP** judgments per role on **relative importance** among [Gf, Gc, WM, Speed, Conscientiousness, Agreeableness, AutonomyPref].
* Convert to a normalized vector **w\_SME**.
* Apply **small deltas** (≤ ±0.2 z) to **means** only if justified; **prefer** to adjust **performance weights** rather than means.
* **Record** AHP matrices and consistency ratio (CR ≤ 0.1 target) per role.

**Bayesian blend (optional):**

μ\_final = λ \* μ\_ONET + (1 - λ) \* μ\_SME

with λ ∈ [0.6, 0.8] initially.

## 7) Deliverables & Schemas

### 7.1 Role Means Table (CSV)

Columns:

role,

Gf\_mean, Gc\_mean, WM\_mean, Speed\_mean,

Conscientiousness\_mean, Agreeableness\_mean, AutonomyPref\_mean,

R\_mean, I\_mean, A\_mean, S\_mean, E\_mean, C\_mean,

source\_ver, onet\_version, mapping\_version

### 7.2 Provenance Table (per role x component)

Columns:

role, engifi\_feature, onet\_component\_name, weight, z\_value, note

Example rows (Backend):

Backend, Gf, Inductive Reasoning (Importance), 0.70, +0.85, primary

Backend, Gf, Deductive Reasoning (Importance), 0.30, +0.60, secondary

...

### 7.3 Autonomy Components Table

Columns:

role, z(Freedom), z(Unstructured), z(Impact), z(Pace\_by\_Equip), z(Automation), AutonomyPref\_mean