Supplementary Table S1. Persona attribute values and conditional dependency constraints used in the MHDS persona generation pipeline.

Attribute	Possible Values	Conditional Constraints
Age	Integer 18–91	*No conditioning.* One of 15 five-year bins (18–22, 22–27, ..., 87–92) is chosen uniformly at random, then an integer is drawn uniformly within that bin. Age subsequently conditions **Education Level**, **Employment Status**, **Marital Status**, and **Number of Children**.
Gender Identity	man, woman, non-binary, agender, genderqueer, transgender	*No conditioning.* Man and woman are each ten times more probable than any of the four gender-minority options, which are equally probable among themselves.
Sexual Orientation	heterosexual, homosexual, bisexual, asexual	*No conditioning.* Heterosexual is strongly dominant; homosexual and bisexual are equally probable but much less frequent; asexual is the least probable. Subsequently conditions **Number of Children**.
City of Living	*Italian:* Rome, Milan, Naples, Bari, Lecce, Bologna *American:* New York, Los Angeles, Kansas City, Philadelphia, Chicago, Washington DC	*No conditioning.* Sampled uniformly from 12 cities (6 Italian, 6 American). Determines the nationality label in **Migration Status**.
Migration Status	native-born Italian, native-born American, immigrant	*Conditioned on **City of Living**.* The native-born status (Italian or American, depending on city) is approximately four times more probable than immigrant, identically for both country contexts.
Religious Beliefs	Christianity, Islam, Hinduism, Buddhism, Judaism, Atheist, Agnostic	*No conditioning.* Atheist is the most probable option, followed by Christianity, Islam, and Hinduism at equal probability, then Agnostic, with Buddhism and Judaism being the least probable and equally rare.

(continued on next page)

(continued from previous page)

Attribute	Possible Values	Conditional Constraints
Education Level	no formal education, primary school, lower secondary, upper secondary, vocational diploma, bachelor's degree, master's degree, PhD	*Conditioned on **Age**.* The eligible set expands with age to reflect plausible attainment trajectories. • **Age = 18**: Upper secondary is strongly dominant. No formal education, primary school, lower secondary, and vocational diploma are equally rare. Bachelor's degree and above are excluded. • **Age 19–21**: Upper secondary remains most probable. Bachelor's degree and vocational diploma are equally likely and moderately common. No formal education and primary school are rare. Lower secondary is excluded. • **Age 22–24**: Bachelor's degree and upper secondary are equally the most probable. Master's degree, vocational diploma, no formal education, and primary school are equally rare. Lower secondary and PhD are excluded. • **Age $\geq$ 25**: All levels are eligible. Upper secondary is modal, followed by vocational diploma, bachelor's degree, lower secondary, master's degree, primary school, with no formal education and PhD being the least probable. Subsequently conditions **Employment Status** and **Occupation**.

(continued on next page)

(continued from previous page)

Attribute	Possible Values	Conditional Constraints
Employment Status	employed full-time, employed part-time, self-employed, full-time student, part-time student, unemployed, retired	*Conditioned on **Age** and **Education Level**.* • **Age < 22, low education** (no formal or primary): Working statuses dominate; full-time student is excluded. • **Age < 22, other education**: Full-time student is strongly dominant over all working statuses. • **Age 22–29, university-level education**: Full-time student and full-time employment are both common, reflecting ongoing postgraduate study alongside early careers. • **Age 22–29, non-university education**: Full-time employment is modal; student statuses are excluded. • **Age 30–39, master's or PhD**: Full-time employment predominates; part-time student status is possible to reflect doctoral continuation. • **Age 30–39, other education**: Full-time employment strongly dominates; student statuses are excluded. • **Age 40–65, low education**: Unemployment is more probable than for higher-educated peers; full-time employment is still the single most likely outcome. • **Age 40–65, higher education**: Full-time employment predominates; a small residual probability for full-time student is retained. • **Age 66–75**: Retired is dominant; part-time employment and self-employment remain possible; unemployment is rare. • **Age > 75**: Retired is near-certain; all working statuses are negligible. Subsequently conditions **Occupation**.

(continued on next page)

(continued from previous page)

Attribute	Possible Values	Conditional Constraints
Occupation	*PhD pool:* university researcher, professor, scientist *University pool:* software dev., data scientist, startup founder, journalist, translator, UX designer, accountant, school teacher, engineer, nurse *Semi-regulated pool:* architect, lawyer, doctor, pharmacist *Non-university pool:* barista, sales associate, electrician, fitness trainer, civil servant, plumber, carpenter, mechanic, retail worker, delivery driver, cleaner, waiter/waitress, hairdresser, chef *Not applicable*	*Conditioned on* ***Education Level*** *and* ***Employment Status****.* • **Unemployed, retired, or full-time student**: set to "not applicable". • **PhD**: PhD-specific roles are the most probable, followed by university-level, then semi-regulated, with non-university roles rarely assigned. • **Master's degree**: PhD-specific roles are excluded. University-level roles are most probable, followed by semi-regulated, then non-university. • **Bachelor's degree**: PhD-specific and semi-regulated roles are excluded. University-level roles are more probable than non-university. • **Other education**: Only the non-university pool is eligible, drawn uniformly from 14 occupations.
Parents' Education	no formal education, primary school, lower secondary, upper secondary, vocational diploma, bachelor's degree, master's degree, PhD	*No conditioning.* Two independent uniform draws from all eight levels, one per parent, unconditioned on the persona's own attributes.

(continued on next page)

(continued from previous page)

Attribute	Possible Values	Conditional Constraints
Marital Status	single, in a relationship, married, divorced, widowed	*Conditioned on **Age**.* • **Age $<$ 22**: Single and in a relationship dominate; married is very rare; divorced and widowed are excluded. • **Age 22–24**: Single remains most probable, followed by in a relationship; married is possible but uncommon; divorced and widowed are excluded. • **Age 25–34**: Single, in a relationship, and married are equally probable; divorced is possible but rare; widowed is excluded. • **Age 35–59**: Married is modal; in a relationship and single are less frequent; divorced and widowed are possible. • **Age 60–69**: Married remains modal; divorced and widowed become more prevalent; single and in a relationship are less common. • **Age $\geq$ 70**: Same as 60–69, with widowed becoming as frequent as divorced. Subsequently conditions **Number of Children**.
Number of Children	Integer: 0, 1, 2, 3, or 4	*Conditioned on **Age**, **Marital Status**, and **Sexual Orientation**.* Zero children is most probable at younger ages and for single personas; higher counts become more probable with age and partnership. • **Single**: Zero children strongly dominates below age 22. One or two children become increasingly possible from age 22 onward; three children are possible from age 30 onward. • **In a relationship**: Zero children dominates below age 22. One child becomes plausible from age 22; from age 30, two children is modal with up to four possible. • **Married**: Zero children remains most likely below age 22. One to two children are most probable at ages 22–29; from age 30, two children is modal with up to four possible. • **Divorced or widowed**: Zero to two children dominate below age 30; from age 30 the distribution converges to the married profile. • **Asexual orientation**: The probability of having no children is increased and the probability of one or more children is reduced, regardless of marital status or age.

(continued on next page)

(continued from previous page)

Attribute	Possible Values	Conditional Constraints
OCEAN Personality Traits	5 traits: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism. Each scored 0–100 and labelled low (<33), moderate (33–66), high (>66).	*No conditioning between traits or with any other attribute.* Each trait is drawn independently and uniformly from integers [0, 100]; the categorical label follows deterministically from the score. All five OCEAN levels are injected into the prompt.
DASS-21 Symptom Severity	For each of depression, anxiety, and stress: severe, moderate, light, without	*No conditioning on any demographic or personality attribute.* Severe, light, and without are equally probable; moderate is slightly less probable. **Rejection rule**: if any one dimension is "severe", the triplet is accepted only if both remaining dimensions are at most "moderate". Triplets pairing "severe" with "light" or "without" on another dimension are discarded and redrawn, preventing implausible co-occurrences. The accepted descriptors are injected into the prompt to calibrate LLM DASS-21 scores and explanations.
Hobbies	hiking, reading novels, baking, football, running, yoga, board games, photography, gardening, volunteering, cinema, classical music, painting, cycling, videogames, travel, cooking, birdwatching, knitting, rock climbing	*No conditioning.* A number $k \in \{2, 3, 4, 5\}$ is drawn uniformly; then k hobbies are sampled without replacement from the pool of 20.
Generation Mode	human-shadow, LLM-assistant	*No conditioning on persona attributes.* Human-shadow in 75% of runs; LLM-assistant in 25%. When LLM-assistant is selected, no persona is constructed and all persona columns remain empty.

Notes. All draws use Python's `random` and `numpy.random` modules. Constraints encode ecological plausibility (e.g., no PhD at age 18; no widowed status below age 25; "severe" on one DASS subscale cannot co-occur with "light" or "without" on another) without targeting any specific national demographic distribution. **Occupation pools** — *PhD-specific:* university researcher, university professor, scientist. *University-level:* software developer, data scientist, startup founder, journalist, translator, UX designer, accountant, school teacher, engineer, nurse. *Semi-regulated:* architect, lawyer, doctor, pharmacist. *Non-university:* barista, sales associate, electrician, fitness trainer, civil servant, plumber, carpenter, mechanic, retail worker, delivery driver, cleaner, waiter/waitress, hairdresser, chef.