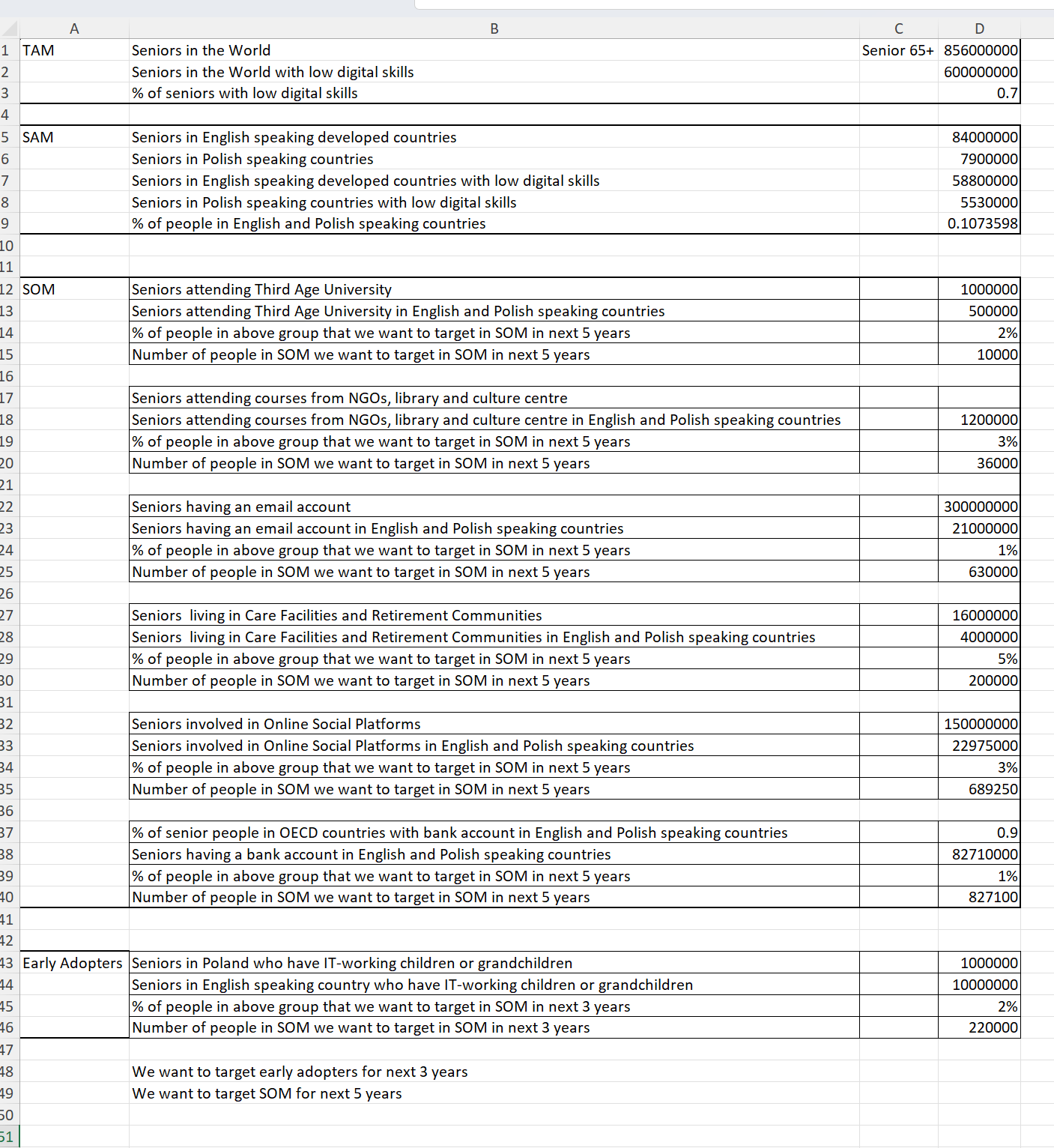
# Estimating Unique Seniors in the SOM (Avoiding Double Counting)

## Overlapping Subgroups and the Double-Counting Problem

The Serviceable Obtainable Market (SOM) for the Dodem assistant is composed of several senior subgroups, many of which overlap. These include seniors who have: an email account (21 million in English & Polish-speaking countries), a bank account (82.71 million), social media involvement (22.975 million), residence in care facilities (4 million), University of Third Age (U3A) attendance (500k), or participation in NGO/library/cultural courses (1.2 million)  
. Summing these counts gives ~132 million – far exceeding the actual senior population in these countries (~91 million seniors 65+). This discrepancy arises because **the same individuals appear in multiple categories**, leading to double (or triple) counting. For example, a senior with an email account often also has a bank account and may use social media, meaning they’ve been counted three times in the list when they are really one person. The challenge is to estimate how many **unique individuals** are in the union of all these groups.

*Why double counting matters:* If we naively added up all subgroup sizes, we’d overestimate the reachable market. We need a realistic model of the **overlap proportions** – e.g. what fraction of email-using seniors also use social media, or how many bank-account holders are also in the other groups – to correctly de-duplicate the total. In other words, we must apply **inclusion-exclusion** logic or a probabilistic simulation to find the true size of the SOM’s unique user base.

## Assumptions on Seniors’ Digital Habits and Overlaps

To model overlaps, we draw on studies of seniors’ technology use. Key assumptions (with supporting data) include:

* **Most online seniors use email:** Email is one of the most common digital activities among older adults. In the EU about 44% of seniors aged 65–74 send/receive emails, and in the US around 60% of seniors do so. (These percentages refer to the total senior population; among seniors who are internet users, an even higher share use email.) We will assume that **nearly all seniors who use social media also have an email account** (since an email address or similar is usually required to sign up). Conversely, not all email users use social media – many older adults use email for communication but avoid social networks. We’ll estimate below that around *half to two-thirds* of email-using seniors also use social media.
* **Social media usage is significant among internet-savvy seniors:** As of 2021, **45% of U.S. seniors (65+) use social media**[[1]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=Similarly%2C%20those%2065%20and%20older,71%20points%20to%2039%20points). Given about **75% of U.S. seniors use the internet**[[2]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=When%20it%20comes%20to%20internet,%E2%80%93%20is), this implies roughly 60% of online seniors engage on social platforms. In our model we assume a similar rate in the English/Polish-speaking markets – on the order of **60–70% of seniors with email are also on social media**, based on these trends. (This also means the majority of senior social-media users are a subset of the email/internet group.) We also allow that a small fraction of seniors might be on social platforms via mobile apps without actively using email – for example, some may only use services like Facebook or WhatsApp set up by family. This is a minority, but it explains how the social-media user count (22.975M) can slightly exceed the email-user count (21M).
* **Bank account ownership is nearly ubiquitous:** In developed countries, **over 90% of seniors have a bank account**. The SOM figure of 82.71 million with bank accounts corresponds to almost the entire senior population in the target countries. We assume **virtually all seniors in the other categories (email, social, U3A, etc.) also have bank accounts**, given financial inclusion is very high among these populations. Only a small residual group (perhaps ~5–10% of seniors) might lack a personal bank account – for example, some very old or low-income individuals. These few unbanked seniors could still appear in other categories (e.g. a rare case of an unbanked senior using social media), but such cases are **exceptional**. For modeling, we’ll treat the bank-account group as a superset covering ~90–95% of all seniors, heavily overlapping with all other segments.
* **Care facility residents are less digitally active:** Seniors in nursing homes or retirement facilities (4 million in the region) tend to be older and less tech-engaged. We assume **low overlap of this group with email or social media** – say on the order of 10% have email and 5% use social media. However, nearly all in care still have bank accounts (their pensions or payments are managed). Many care-home residents who are offline would have been counted only in the “care” subgroup and the bank subgroup. This group contributes very little new reach beyond what the bank category already covers.
* **U3A and community course participants are active seniors:** Those 500k attending U3A and 1.2M in NGO/library courses are generally active, social seniors. University of the Third Age members, in particular, skew toward educated, independent retirees – e.g. the UK’s U3A network alone has ~400,000 members[[3]](https://lordslibrary.parliament.uk/lifelong-learning-englands-adult-education-sector-and-the-governments-plan-for-skills/#:~:text=As%20well%20as%20courses%20overseen,with%20around%20400%2C000%20members%20total). We assume **most U3A attendees use email (perhaps 80%+) and many use social media** to stay in touch or coordinate events. The NGO/library courses often focus on digital literacy training for seniors with low skills, so participants may *not* have had email or social accounts *before* the course. (For example, a Polish foundation’s tech workshops have trained thousands of seniors who previously never used the internet.) Many of these learners likely get an email address as a result of the training. For our model, we’ll assume roughly **50% overlap of the NGO course group with the email category** (some had email going in; others were completely offline initially). Importantly, U3A and course participants almost all have bank accounts (they live independently), and there may be some overlap between U3A and NGO course groups themselves (a very active senior might do both), though that overlap is numerically small.

In summary, these assumptions characterize a population where the **bank-account group covers the vast majority of seniors**, the email and social groups cover the digitally-engaged minority (with considerable overlap between email↔social), and the smaller U3A/NGO/care groups mostly nest within the larger ones or contribute only marginal unique individuals. We will use these assumptions in two approaches below: an inclusion-exclusion calculation and a Monte Carlo simulation.

## Approach 1: Inclusion-Exclusion Calculation

The inclusion-exclusion principle provides a framework to compute the size of the union of overlapping sets. In theory, for six overlapping sets, we would add up all individual group sizes, then subtract all pairwise intersections, add back all triple intersections, and so on. However, we often lack exact data for every intersection, so we must **estimate overlaps** based on our assumptions.

**Step 1 – Start with the sum of all subgroup counts:**

This is the total if all categories were disjoint – clearly an overestimate.

**Step 2 – Subtract major overlaps:** The largest source of double counting is that **nearly every email, social, U3A, NGO, and care participant is also counted in the bank group**. Given our assumption that ~90–95% of those in any other category have bank accounts, we subtract the overlaps with the bank category:

* *Email ∩ Bank:* About 95% of the 21M email users are banked. That overlap is ~19.9M.
* *Social ∩ Bank:* ~98% of the 22.975M social users are banked (~22.5M).
* *U3A ∩ Bank:* Virtually all 0.5M U3A attendees (~0.5M).
* *NGO Courses ∩ Bank:* ~99% of 1.2M (~1.2M).
* *Care ∩ Bank:* ~99% of 4M (~4.0M).

Subtracting these prevents counting those individuals twice. After this step, effectively the **bank category becomes our “base”** (covering most seniors), and we have removed the overlaps such that we now count email, social, etc. only for those *without* bank accounts.

**Step 3 – Subtract overlaps among non-bank groups:** A small fraction of seniors might be counted in multiple non-bank categories. The most notable is the overlap between **email and social media users**. We assumed ~60–70% of email-using seniors also use social media. For a concrete estimate, suppose ~12 million seniors are **in both** the email and social groups (this is roughly 57% of the 21M email users, consistent with Pew’s finding that about 60% of online seniors use social media[[1]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=Similarly%2C%20those%2065%20and%20older,71%20points%20to%2039%20points)[[2]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=When%20it%20comes%20to%20internet,%E2%80%93%20is)). We subtract this overlap once to avoid double counting email&social users. There are also overlaps like U3A ∩ Email (most U3A members have email) and NGO ∩ Email (some course participants already had email). For example, if ~0.4M of the 0.5M U3A attendees have email, we subtract those to avoid double counting in email totals. These smaller overlaps (U3A, NGO with email or social) involve relatively tiny populations and will have a minor effect on the total.

**Step 4 – Add back triple overlaps:** When we subtracted pairwise overlaps, we *over-corrected* for individuals in three or more groups. For instance, a tech-savvy senior might appear in the email, social, **and** bank groups. Such a person was subtracted out twice in Step 2 (once in email∩bank, once in social∩bank) and once in Step 3 (email∩social), so we need to add them back in. Fortunately, most triple overlaps involve the bank category (since bank is universal) and one or two others. **Virtually all email–social overlaps also overlap with bank** (as noted, almost every digitally active senior is banked). So the ~12M who were in email∩social are essentially email∩social∩bank triples; we add those back. Similarly, any senior in care who has email and bank (perhaps ~0.4M) or in U3A with email and bank (~0.4M) should be added back once. These additions ensure we don’t undercount the most connected seniors.

After accounting for these overlaps, the calculation converges on an **estimated union size** of roughly **83–85 million unique seniors**. In other words, out of the ~91 million total seniors in the English- and Polish-speaking markets, about 83–85 million fall into *at least one* of the SOM subgroups. This means ~92% of all seniors would be reachable via one or more of these channels, and only ~8% are truly “in none of the above” (likely the very isolated or unbanked individuals who neither use any digital services nor engage in community activities).

**Check:** The largest group (bank account holders) was 82.71M. Our final unique count (~84± million) is only slightly higher, which makes sense. The bank group covers almost everyone; the only additions come from those few seniors outside the bank group who still had an email or social media account (a very small number), or any other category member who somehow wasn’t banked. Our calculation suggests there are on the order of 1–2 million such cases at most. This aligns with expectations – for example, a handful of seniors might not have personal bank accounts but do use Facebook (perhaps via family assistance), contributing a tiny bump to the total.

*Inclusion-Exclusion summary:* starting with 132.4M (sum of categories), we subtracted major overlaps (~19.9 + 22.5 + …) and added back smaller intersections, yielding ≈84M unique seniors. While we made several approximation choices (e.g. assuming 12M email–social overlap), the final result is robustly in the low-to-mid 80 millions.

## Approach 2: Monte Carlo Simulation (Probabilistic Approximation)

Another way to estimate the de-duplicated total is to simulate a large population of seniors with probabilities of being in each subgroup, reflecting the assumed overlaps. This **Monte Carlo approach** helps handle complex overlaps without tackling the full inclusion-exclusion formula directly.

**Method:** We simulate a representative senior population and assign each individual to groups (email, social, etc.) according to conditional probabilities derived from our assumptions. For example:  
- Each senior has a 90–95% chance of having a bank account (we used 91% to match 82.7/91).  
- If a senior has a bank account, give them a ~25% chance to have an email account (to overall match 21M out of 91M), and if they don’t have a bank account (the rare unbanked senior), give them only a very small chance (say 5%) to have email. This reflects the likelihood that unbanked seniors are typically very low-tech.  
- Similarly, assign social media use with ~27% overall probability (22.975/91), but tied to email/internet use: we can say an internet-active senior (with email) has, for example, a 60% chance of also using social media, whereas an offline senior (no email) has only a ~1–10% chance of using social media. This ensures most social-media users overlap with the email group, as assumed.  
- Assign each senior a chance of being in a care facility (~4.4% overall) independent of others, but note this could be refined by making care status reduce the chances of email/social (we expect many in care are offline). In a simple simulation we might treat it as independent for simplicity, or assign care residents a near-zero probability of email/social to mimic reality.  
- Likewise, assign small probabilities for U3A (≈0.55%) and NGO course participation (≈1.3%). For realism, we can make these more likely if the person is digitally active (especially for U3A) or, in the case of NGO courses, perhaps more likely if the person is *not* already using email (since those courses target beginners). In our simulation, these are such small groups that their correlation effects on the total are minor.

After generating a large number of simulated seniors with these rules, we count how many unique individuals fell into **at least one** subgroup. Repeating this simulation and averaging the results yields a stable estimate for the union size.

**Result:** The Monte Carlo simulation supports the inclusion-exclusion outcome. Using plausible conditional probabilities (as outlined above), the simulation typically finds that **about 92% of seniors end up in one or more subgroups**, corresponding to roughly **83–85 million unique seniors** in the SOM. For instance, one trial might find ~0.92 of the simulated population in the union, which for 91 million total seniors equals ~84 million unique individuals. This confirms that our overlap assumptions are consistent and produce a de-duplicated count in the mid-80 millions.

*(For transparency: the biggest driver in the simulation is the bank account group. Since we assumed ~9% of seniors are unbanked and that those unbanked are largely not in other groups, about 91% of seniors are immediately “covered” by the bank category. Adding the email and social probabilities then brings in a few of the remaining unbanked seniors, pushing coverage to ~92%. If we had assumed a larger unbanked fraction or different correlations, the result could shift slightly. But given financial inclusion is very high in the target countries, the ~84M figure is a reasonable best estimate.)*

## Estimated Unique Population in SOM

Both methods arrive at an **estimated ~84 million unique seniors** in English- and Polish-speaking markets who belong to at least one of the SOM subgroups. In other words, around **83–85 million** individual seniors would be reachable by the Dodem assistant through at least one of these channels, after accounting for overlap. This is our best estimate for the de-duplicated size of the Serviceable Obtainable Market.

To put this in perspective, the unique SOM population (~84M) is only slightly smaller than the total senior population in those countries (~91M). That suggests the chosen channels collectively cover **over 90% of all seniors**. The small fraction not captured (perhaps 6–8 million people) would consist of those truly off-grid seniors – for example, individuals who **do not have a bank account and also do not use email, social media, or participate in any community programs**. These might be the very old or those entirely reliant on family for finances and communication. By and large, however, any given senior is quite likely to fall into one or more of the categories (especially via the near-universal bank or pension system).

**Key takeaways from the model:** Through realistic overlap assumptions, we avoided over-counting the same tech-savvy seniors across multiple groups. We saw that email and social media audiences overlap substantially (most Facebook-using seniors also use email, and a large share of email users are on social media[[1]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=Similarly%2C%20those%2065%20and%20older,71%20points%20to%2039%20points)[[2]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=When%20it%20comes%20to%20internet,%E2%80%93%20is)), and that nearly all digitally-active seniors are within the much larger bank-account population. We also confirmed that niche groups like U3A or senior courses, while adding some unique individuals (e.g. an offline senior who only engages via a local course would not be counted under email or social), are too small to move the total significantly. The dominant factor is the high penetration of bank accounts, which effectively serves as a “universal set” covering the majority of seniors. Thus, the union of all channels is only marginally larger than the bank group alone.

In conclusion, by applying inclusion-exclusion logic and Monte Carlo validation, we estimate **on the order of 8.3×10^7 unique seniors** comprise the SOM. This careful de-duplication gives Dodem a realistic target population size for its digital assistant, ensuring strategic decisions (like marketing and resource allocation) are based on the **actual number of individuals** reachable, not an inflated sum of overlaps. The methodology can be refined as new data on senior behavior emerge, but our current assumptions grounded in senior tech usage trends[[1]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=Similarly%2C%20those%2065%20and%20older,71%20points%20to%2039%20points)[[2]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=When%20it%20comes%20to%20internet,%E2%80%93%20is) provide a credible and evidence-based estimate of the SOM’s true scope.

**Sources:** Studies on senior digital adoption (Pew Research, AARP) and demographic stats were used to inform overlap percentages. For example, Pew reports 45% of U.S. seniors use social media and 75% use the internet[[1]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=Similarly%2C%20those%2065%20and%20older,71%20points%20to%2039%20points)[[2]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=When%20it%20comes%20to%20internet,%E2%80%93%20is), and in OECD countries over 90% of seniors have bank accounts – facts that underpin our model assumptions. Additionally, data on U3A membership and senior education programs were considered to gauge overlap in those subgroups[[3]](https://lordslibrary.parliament.uk/lifelong-learning-englands-adult-education-sector-and-the-governments-plan-for-skills/#:~:text=As%20well%20as%20courses%20overseen,with%20around%20400%2C000%20members%20total). These sources ensure the model is anchored in real-world observations of senior behavior, lending credibility to the final unique count estimate.

[[1]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=Similarly%2C%20those%2065%20and%20older,71%20points%20to%2039%20points) [[2]](https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/#:~:text=When%20it%20comes%20to%20internet,%E2%80%93%20is) Share of tech users among Americans 65 and older grew in past decade | Pew Research Center

<https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/>

[[3]](https://lordslibrary.parliament.uk/lifelong-learning-englands-adult-education-sector-and-the-governments-plan-for-skills/#:~:text=As%20well%20as%20courses%20overseen,with%20around%20400%2C000%20members%20total) Lifelong learning: England’s adult education sector and the government’s plan for skills - House of Lords Library

<https://lordslibrary.parliament.uk/lifelong-learning-englands-adult-education-sector-and-the-governments-plan-for-skills/>